

The background of the book cover is a detailed, blue-tinted image of a microchip or printed circuit board. It shows various electronic components, traces, and a grid-like pattern, suggesting a high-tech or digital theme.

**Second Edition**

Introduction to  
**DIGITAL MOBILE  
COMMUNICATION**

**Yoshihiko Akaiwa**

**WILEY**

# **INTRODUCTION TO DIGITAL MOBILE COMMUNICATION**

# **INTRODUCTION TO DIGITAL MOBILE COMMUNICATION**

---

Second Edition

**YOSHIHIKO AKAIWA**

**WILEY**

Copyright © 2015 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at [www.copyright.com](http://www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at [www.wiley.com](http://www.wiley.com).

***Library of Congress Cataloging-in-Publication Data:***

Akaiwa, Yoshihiko.

Introduction to digital mobile communication / Yoshihiko Akaiwa. – Second edition.

Includes bibliographical references and index.

ISBN 978-1-119-04110-8 (hardback)

1. Mobile communication systems. 2. Digital communications. I. Title.

TK6570.M6A39 2015

621.3845'6–dc23

2015005234

Cover image courtesy of the Getty Images © Inok.

Set in 10/12pt Times by SPi Global, Pondicherry, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1



# CONTENTS

<b>Preface to the Second Edition</b>	<b>xiii</b>
<b>Preface to the First Edition</b>	<b>xv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Digital Mobile Radio Communication System	1
1.2 The Purpose of Digitization of Mobile Radio Communications	5
1.2.1 Data Communication	5
1.2.2 Voice Scrambling	6
1.2.3 Spectrum Efficiency	6
1.2.4 System Cost	7
<b>2 Signal and Systems</b>	<b>9</b>
2.1 Signal Analysis	9
2.1.1 Delta Function	9
2.1.2 Fourier Analysis	15
2.1.3 Signals	26
2.1.4 Digital Signals	31
2.1.5 Modulated Signals	34
2.1.6 The Equivalent Base-Band Complex Expression	36
2.2 Noise Analysis	37
2.2.1 Noise in Communication System	37
2.2.2 Statistics of Noise	39
2.2.3 Power Spectral Density of Noise	42

2.2.4	Autocorrelation Function of Filtered Noise	43
2.2.5	Bandpass Noise	44
2.2.6	Envelope and Phase of a Sinusoidal Signal in Bandpass Noise	48
2.2.7	Generation of Correlated Noises and its Probability Density Function	49
2.2.8	Sums of Random Variables and the Central Limit Theorem	51
2.3	Linear System	55
2.3.1	Linear Time-Invariant System	55
2.3.2	Response of Linear System	55
2.3.3	System Description with Differential Equations	63
2.3.4	Examples of Linear Systems	66
2.4	Discrete-time System	75
2.4.1	Sampling and the Sampling Theorem	75
2.4.2	The Energy, Power, and Correlation of Discrete-Time Signals	78
2.4.3	The Fourier Transform of Discrete-Time Signals	79
2.4.4	Response of Discrete-Time System	85
2.4.5	Description with Difference Equation	92
2.4.6	Digital Filter	94
2.4.7	Downsampling, Upsampling, and Subsampling	98
2.4.8	Inverse Circuit	101
2.4.9	Window Function	101
2.4.10	Discrete Fourier Transform	102
2.4.11	The Fast Fourier Transform	106
2.5	Optimization and Adaptive Signal Processing	108
2.5.1	Solution of Optimization Problem	108
2.5.2	Adaptive Signal Processing	112
Appendix 2.A	$\lim_{\Omega \rightarrow \infty} (\sin \Omega t / \pi t) = \delta(t)$	124
Appendix 2.B	Conditions for a Test Function for the Delta Function, $\lim_{T, \Omega \rightarrow \infty} \int_0^T g(t) \sin \Omega t dt = 0$	125
Appendix 2.C	Formulae for the Trigonometric Functions	126
References		126
<b>3</b>	<b>The Elements of Digital Communication System</b>	<b>127</b>
3.1	Pulse Shaping	127
3.1.1	Nyquist's First Criterion	128
3.1.2	Nyquist's Second Criterion	132
3.1.3	Nyquist's Third Criterion	134
3.1.4	Other Pulse-Shaping Methods	135
3.2	Line Coding	137
3.2.1	Unipolar (On–Off) Code and Polar Codes	137

3.2.2	Multilevel Codes	137
3.2.3	The Gray Codes	138
3.2.4	Manchester (Split-Phase) Code	139
3.2.5	Synchronized Frequency Shift Keying Code	141
3.2.6	Correlative Coding	141
3.2.7	Differential Encoding	148
3.3	Signal Detection	149
3.3.1	$C/N$ , $S/N$ , and $E_b/N_0$	149
3.3.2	Bit Error Rate	150
3.3.3	NRZ Signaling with Integrate-and-Dump Filter Detection	156
3.3.4	Nyquist-I Signaling System	157
3.3.5	The Matched Filter	157
3.3.6	Joint Optimization of the Transmit and the Receive Filters	162
3.3.7	The Optimum Receiver	164
3.3.8	The Maximum-Likelihood Receiver and the Viterbi Algorithm	170
3.3.9	The Optimum Receiver for Signals without Intersymbol Interference	174
3.4	Synchronization	175
3.4.1	Symbol Timing Recovery	175
3.4.2	Frame Synchronization	176
3.5	Scrambling	177
3.6	Public Key Cryptosystem	180
3.7	Multiplexing and Multiple Access	182
3.8	The Channel Capacity	183
Appendix 3.A	Fermat's Theorem and the Chinese Remainder Theorem	185
References		187
<b>4</b>	<b>Mobile Radio Channels</b>	<b>189</b>
4.1	Path Loss	190
4.2	Shadowing	193
4.3	Fast Fading	193
4.3.1	RF Power Spectrum Spread due to Fast Fading	195
4.3.2	Correlations Between the In-phase and Quadrature Components	196
4.3.3	Correlation of the Envelope	197
4.3.4	Spatial Correlation of the Envelope	198
4.3.5	Random Frequency Modulation	198
4.4	Delay Spread and Frequency-Selective Fading	200
4.4.1	Coherence Bandwidth	202
4.4.2	Frequency-Selective Fading	203

4.5	The Near–Far Problem	204
4.6	Cochannel Interference	205
4.6.1	Rayleigh Fading	206
4.6.2	Shadowing	206
4.6.3	Combined Fading and Shadowing	207
4.6.4	Discussion	207
4.7	Receive Power Distribution and Radio Channel Design	207
4.7.1	Receive Power Distribution	209
4.7.2	Channel Link Design	210
Appendix 4.A	Propagation Loss Formula	214
Appendix 4.B	Interference Probability under Shadowing	216
Appendix 4.C	Interference Probability under Combined Fading and Shadowing	217
	References	217
<b>5</b>	<b>Elements of Digital Modulation</b>	<b>219</b>
5.1	Digitally Modulated Signals	219
5.2	Linear Modulation Versus Constant Envelope Modulation	220
5.3	Digital Modulations	221
5.3.1	Phase Shift Keying	221
5.3.2	Frequency Shift Keying	226
5.3.3	Constant Envelope PSK	228
5.3.4	Quadrature Amplitude Modulation	229
5.4	Power Spectral Density of Digitally Modulated Signals	229
5.4.1	Linear Modulation	231
5.4.2	Digital FM	231
5.5	Demodulation	233
5.5.1	Coherent Detection	233
5.5.2	Envelope Detection	245
5.5.3	Differential Detection	246
5.5.4	Frequency Discriminator Detection	250
5.5.5	Error Rates in Fading Channels	264
5.6	Computer Simulation of Transmission Systems	270
Appendix 5.A	Distortion of Modulated Signal Applied to a Nonlinear Circuit	275
Appendix 5.B	Derivation of the Expected Gaussian Noise Power for Frequency Discriminator	276
Appendix 5.C	M–Sequence Generator	277
	References	278
<b>6</b>	<b>Digital Modulation/Demodulation for Mobile Radio Communication</b>	<b>281</b>
6.1	Digital Modulation for Analog FM Mobile Radio Systems	282

6.2	Constant Envelope Modulation	282
6.2.1	MSK	283
6.2.2	Partial-Response Digital FM	294
6.2.3	Nyquist-Filtered Digital FM	306
6.2.4	Performance Comparison	310
6.3	Linear Modulation	313
6.3.1	$\pi/4$ -Shifted QPSK	315
6.3.2	Eight-Level PSK	320
6.3.3	16QAM	322
6.4	Spread-Spectrum System	322
6.5	Multicarrier Transmission	329
6.5.1	Orthogonal Frequency-Division Multiplexing	329
6.5.2	Generation of Multicarrier Digital Signal	337
6.5.3	Demodulation of Multicarrier Signals	341
6.6	Single-Carrier Frequency-Division Modulation	343
Appendix 6.A	Mathematical Principles of Orthogonal Frequency-Division Multiplexing	346
	6.A.1 Band-Limited System	347
	6.A.2 Nonband-Limited System	348
	References	349
<b>7</b>	<b>Other Topics in Digital Mobile Radio Transmission</b>	<b>355</b>
7.1	Diversity Transmission System	355
7.1.1	Probability Density Function of SNR for Diversity System	357
7.1.2	Average Error Rate for Diversity Systems	360
7.1.3	Multiple Transmitter Diversity System	367
7.1.4	Antenna Selection Diversity System	370
7.2	Multi-Input Multi-Output Systems	375
7.2.1	Maximal Ratio Combining Diversity Systems	375
7.2.2	Space-Time Codes	385
7.2.3	SDM in MIMO Systems	386
7.3	Adaptive Automatic Equalizer	401
7.3.1	Linear Equalizer	402
7.3.2	Performance Criteria for Equalization	405
7.3.3	Decision Feedback Equalizer	409
7.3.4	The Viterbi Equalizer	410
7.3.5	Adaptation and Prediction Algorithm	411
7.3.6	Preequalization	411
7.3.7	Frequency-Domain Equalizer	418
7.3.8	Turbo Equalizer	419
7.3.9	Discussions on Equalization	419
7.3.10	Applications to a Mobile Radio Channel	421

7.4	Error Control Techniques	422
7.4.1	Linear Block Codes	424
7.4.2	Cyclic Codes	426
7.4.3	Convolutional Codes	429
7.4.4	Concatenated Codes	430
7.4.5	Turbo Codes	430
7.4.6	LDPC Code	444
7.4.7	A Phenomenological Expression of the a Priori Probability and Error Rates	449
7.4.8	ARQ	452
7.4.9	Applications to Mobile Radio Channels	453
7.5	Trellis-Coded Modulation	453
7.6	Adaptive Interference Cancellation	456
7.6.1	Adaptive Array Antenna	457
7.6.2	Adaptive Interference Suppression	466
7.6.3	Discussion	467
7.7	Voice Coding	469
7.7.1	Pulse Code Modulation	470
7.7.2	Delta Modulation	471
7.7.3	Adaptive Differential Pulse Code Modulation	472
7.7.4	Adaptive Predictive Coding	473
7.7.5	Multipulse Coding	476
7.7.6	Code-Excited Linear Predictive (CELP) Coding	477
7.7.7	LPC Vocoder	482
7.7.8	Application to Mobile Radio Communications	482
Appendix 7.A	Average Error Rate for Maximal Ratio Combiner with Coherent Detector	484
Appendix 7.B	Average Error Rate of Maximal Ratio Combining System with Coherent Detector with Use of Approximate Probability Density Function	485
	References	486
<b>8</b>	<b>Equipment and Circuits for Digital Mobile Radio</b>	<b>493</b>
8.1	Base Station	493
8.2	Mobile Station	494
8.3	Superheterodyne and Direct Conversion Receivers	495
8.3.1	Image Rejection Downconverter	497
8.4	Transmit and Receive Duplexing	501
8.5	Frequency Synthesizer	501
8.6	Transmitter Circuits	503
8.6.1	Digital Signal Waveform Generator	503
8.6.2	Modulator	504
8.6.3	Linear Power Amplifier	507

8.6.4	Transmit Power Control	525
8.7	Receiver Circuits	527
8.7.1	AGC Circuit	527
8.7.2	Signal Processing with Logic Circuits	529
8.7.3	Demodulator	532
8.8	Countermeasures Against dc Blocking and dc Offset	535
Appendix 8.A	Quarter-wavelength Line	538
References		539
<b>9</b>	<b>Digital Mobile Radio Communication Systems</b>	<b>543</b>
9.1	Fundamental Concepts	543
9.1.1	The Cellular Concept	543
9.1.2	Multiple Access	551
9.1.3	Channel Assignment	554
9.1.4	Multiple-Access System	563
9.1.5	Intercell Interference Suppression	566
9.1.6	Repeater System	566
9.1.7	A Performance Analysis of Digital Cellular System	567
9.2	Digital Transmission in Analog Mobile Communication Systems	577
9.3	Paging Systems	578
9.4	Two-Way Digital Mobile Radio	579
9.5	Mobile Data Service Systems	580
9.5.1	MOBITEX	580
9.5.2	Teleterminal System	580
9.5.3	Mobile Data Systems in Analog Cellular Systems	580
9.6	Digital Cordless Telephone	581
9.6.1	Second-Generation Cordless Telephone	581
9.6.2	Digital European Cordless Telecommunications	582
9.6.3	Personal Handy System	582
9.7	Digital Mobile Telephone Systems	583
9.7.1	The GSM System	584
9.7.2	Digital Cellular Systems in North America	587
9.7.3	Digital Cellular Systems in Japan	591
9.7.4	Evolution of the Second-Generation Systems	592
9.7.5	The Third-Generation System	592
9.7.6	Evolution of 3G Systems	595
9.7.7	WiMAX	599
9.7.8	The Fourth-Generation System	600
9.8	Wireless Local Area Network	600
9.8.1	IEEE 802.11 Series	600
9.8.2	Bluetooth	605
9.8.3	UWB	605

9.8.4	ZigBee	606
9.8.5	BWN	606
9.8.6	MBWA	608
Appendix 9.A	Poisson Arrival Rates	608
References		609
<b>Index</b>		<b>613</b>



# **PREFACE TO THE SECOND EDITION**

It has been 17 years since the publication of this book in 1997. The progress of technology and penetration of services in the digital mobile communications field, especially in cellular systems, is much faster than I expected. In general, the mobile radio communication services are widely accepted as the wireline internet services of the world. My intention of rewriting this book is to provide an update on new technologies. However, technology is evolving so rapidly that it is difficult to fulfill this. Therefore, this book might need to be updated regularly to be in pace with the technology.

The central motive force for technological progress in the mobile radio field has been the digitization of communications to provide digital communication services including digital voice, message, and mobile internet (data) services. In physical layer technologies for cellular systems, digital modulation and multiplexing schemes abandoned the use of constant envelope signals, since spectrum efficiency with non-constant envelope signals is given a higher priority than power efficiency with the constant envelope signals. In this scenario, linear modulations, code-division multiplexing, and the latest orthogonal frequency multiplexing are developed. To make these techniques more feasible, radio frequency (RF) devices with a nonlinearity compensation technique to get simultaneously a high linearity and power efficiency appeared. Together with the high spectrum efficiency, a higher transmission rate is pursued for low-latency services. High-level digital modulations such as 16- and 64QAM, multi-antenna (multi-input multi-output (MIMO)) systems, and fast scheduling schemes of radio packet transmission have been developed. The other newly appeared ones are turbo codes, LDPC code, G. 729 voice coding, frequency domain equalization, and digital predistortion amplifiers.

The evolution of these technologies has been supported with amazing progress of semiconductor circuits, which made the complex signal processing feasible.

Many books have been published about the progress of new topics in mobile radio communications. The purpose of this book is to give a comprehensive introduction to digital mobile communications. I tried my best to include every important technology in digital mobile communications, spanning from physical layer technologies such as circuits, devices, pulse transmission, digital modulations/demodulations, signal multiplexing, voice coding, multiantenna systems, and mobile radio channels to higher layer techniques including error control, channel assignments, and cellular system design. Although many topics became obsolete, I still have included some of them in this version, since I think those topics will be useful to recall and recognize the technological progress in the field.

While I try to understand new technologies, I recognize the importance of their principles and the fundamental knowledge of treating signals and systems. Therefore, I have described the physical meanings in short together with the use of mathematical models and expanded the description on signals and systems in Chapter 2.

Personally, I am happy to be able to be involved with digital mobile communications and watch its progress from the very initial stage till now at my professional retirement. I hope this revised book with updated descriptions, which reflects my experience in the field, will be helpful to readers.

I would like to thank the people who helped me to revise this book: Associate Prof. Osamu Muta for providing some results as figures and useful comments on the manuscript; Profs. Fumiyuki Adachi, Mamoru Sawahashi, Takeo Ohgane, and Tetsuki Taniguchi for helpful discussions on new technologies; and Kouji Iwaki and Gakulin Ma for calculation and for preparing some figures.

# **PREFACE TO THE FIRST EDITION**

Digitization is a technical trend in telecommunications as well as in other fields, for example, digital audio systems (compact disk) and digital control of machines. A digital approach enjoys higher accuracy and stability of the system over an analog system. Driving forces for digitization are VLSI (very large scale integrated) circuits and computers, which make it feasible to implement circuits required for the digitization of systems. In the telecommunications field, technological advances have been made on data and digital speech transmission on switched telephone networks, digital microwave communications, and digital fiber-optic communications. Digitization in mobile radio communications has lagged behind in comparison with these developments. However, in recent times, explosive activity in research and development of digital mobile communications seems yet to recover the delay in the digitization progress. This progress is further spurred by novel applications of digital mobile communications such as personal communication services, mobile computing, and mobile multimedia.

Some digital technologies can be applied in common to any field. However, some are not directly applicable because of differences in the requirements of a specific technical field. For mobile communications, robustness against fast fading, spectrum efficiency, power efficiency, and compactness and low price of equipment are essential. Digital modulation/demodulation is the key technology to fulfill these requirements for mobile communications.

This book is intended to introduce the digital mobile communication technologies with an emphasis on digital transmission methods.

Chapter 1 introduces what a digital mobile communication system is and why mobile communications are digitized. The reader will get a rough global insight into the various technologies used in digital mobile communications.

Chapter 2 describes mathematical analyses of signals and noise. Signal analysis includes topics such as the delta function, the Fourier transform, the response of a linear system, the impulse response, the cross correlation, the autocorrelation, a representation of digital signals, the average signal power, the power spectral density, modulated signals, orthogonal signals, and the sampling theorem. The analysis of noise includes the noise figure or the noise temperature and the statistical characteristics of the noise.

Chapter 3 covers the fundamentals of a digital transmission system. It deals with the Nyquist criteria for transmission without intersymbol interference, multilevel coding, partial response systems, a matched filter receiver, and the optimum receiver.

Chapter 4 covers mobile radio channels. It includes some specific features of mobile radio channels, such as Rayleigh fading, frequency-selective fading, shadow fading, and the near-far problem.

Chapter 5 is devoted to the elements of digital modulation. The chapter includes a fundamental description of digital modulation, the power spectrum of modulated signals, demodulation, and error rate analysis.

Chapter 6 discusses digital modulation methods for mobile radio communications. Constant envelope digital modulations are described first. These are minimum shift keying, tamed frequency modulation, GMSK (Gaussian MSK), Nyquist multilevel FM, PLL-QPSK, CCPSK, duoquaternary FM, correlative PSK, and digital PM. Then linear digital modulations are described. Following the description of the significance and the problems of linear modulation when applying it to mobile radio communications, some digital modulations such as  $\pi/4$ -shifted QPSK, QPSK, and 16QAM are discussed. For the circuit aspects related to linear modulations, techniques for the nonlinearity compensation in a power amplifier are introduced.

Chapter 7 describes other topics related to digital mobile radio communication. It includes antipath modulations, multicarrier systems, spread spectrum communications, diversity communication systems, adaptive equalizers, error control techniques, trellis-coded modulation, adaptive interference cancellations, and voice coding.

Chapter 8 describes equipment and circuit implementation methods. Configurations of base stations and mobile stations are shown. It also includes a discussion on superheterodyne and direct conversion receivers, transmit and receive duplexing, frequency synthesizers, transmitter circuits, receiver circuits, and countermeasures against dc blocking and dc offset.

Chapter 9 describes digital mobile radio systems. Fundamental concepts such as traffic theory, the cellular concept, multiple access systems, channel assignment, and interbase-station synchronization are discussed first. Then systems such as digital transmission in analog FM systems, paging systems, the so-called two-way digital mobile radio, mobile data service systems, digital cordless phone, digital mobile telephone systems, wireless local area networks, mobile satellite systems, personal communication, and FPLMTS are introduced.

The author's objectives in writing this book are as follows:

- Almost all important topics are covered, although the depth of their description differs.
- A beginner engineer can follow the book by starting from studying the analyses of signals and noise and basic technologies for digital communications.
- Intuitive understanding of the meaning of the techniques is emphasized rather than rigor of the discussion.
- Circuit implementation methods are described that are useful for design engineers of digital mobile communication systems.

A certain part of this book is bound to become obsolete in the near future, when more advanced technologies on digital mobile communications are developed as the result of current ongoing active research in the field. However, the author believes that this book may be helpful to mobile radio engineers, since there is no book that comprehensively describes digital mobile communications at this time.

Engineering is a little different from science. In contrast to science, new technologies are sometimes achieved through a combination of well-known facts or by introduction of a technology that was developed in another field. It is rather easy to understand the mechanism, the importance, and the motivation of a new technique once it has been presented. But it is fairly difficult for engineers to create a new technology. Profound understanding of fundamental key technologies in conjunction with assumed background is important. The most important things for creative engineers as well as scientists are motivation for and patience with their technological goal. I hope this book will help mobile radio engineers in their work in this new field of technology.

## ACKNOWLEDGMENTS

It is my sincere pleasure to thank individuals who helped me to prepare this book. Some of the contents are based on my work at NEC corporation. I thank Messrs. Ichiroh Takase, Yoshinori Nagata, and Yukitsuna Furuya for their cooperation. Since I joined the university, I have enjoyed working with my research assistant and students. Some results of the research carried out with them at the university are also included in this book. Furthermore, they helped me by preparing and reviewing the manuscript. I wish to thank all of them, especially assistant professor Hiroshi Furukawa, Tetuhiko Miyatani, Hidehiro Andoh, Toshio Nomura, Taisuke Konishi, Masahiro Maki, Mokoto Taromaru, Takayuki Torigai, Kuninori Oosaki, Hisao Koga, Canchi Radhakrishna, Fangwei Tong, Masaya Saitou, Hitoshi Iochi, Mutsuhiko Ooishi, Takenobu Arima, Takashi Matubara, Takuya Ootani, and Takashi Seto.

I would also like to thank Mrs. Kara Maria and my wife Terumi for typing and proofreading some parts of the manuscript. My special thanks go to Professor Jeffrey Capone of the Northeastern University for his generous reviewing of the manuscript and correction of the English language. I am very grateful to the series editor Professor John G. Proakis of the Northeastern University and the associate publisher (Retired) Mr. George Telecki of John Wiley & Sons for their efforts to publish this book.

---

# 1

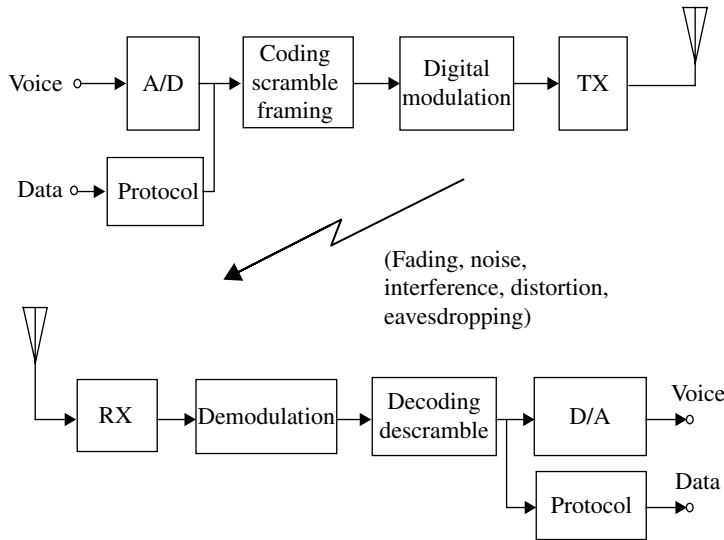
---

## INTRODUCTION

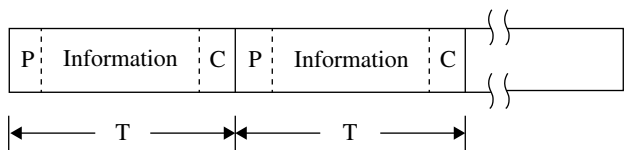
This chapter briefly describes the digital mobile radio communication system and the significance of digitization in mobile radio communications.

### 1.1 DIGITAL MOBILE RADIO COMMUNICATION SYSTEM

A schematic block diagram of a digital mobile radio communication system is shown in Figure 1.1. A voice signal is converted into a digital signal via a voice coder. The digitized voice signal is transmitted through a digital mobile radio channel and is converted into a reconstructed analog voice signal at the receiver. The target of voice coding technology is to achieve a lower coding rate while keeping an acceptable voice quality. The digital signals are processed by logic circuits for several system requirements: typically, channel coding, scrambling, and framing. Channel coding is a process which inserts additional bits in order to correct or to detect errors in received signals. Scrambling is designed to hide a transmitted signal from a third party by performing a complicated transformation which is known only to an authorized recipient. Framing is a process where information signals are grouped into blocks with other signals as shown in Figure 1.2. The purpose of framing is to multiplex different signals in the time domain (time division multiplexing (TDM)) to introduce channel coding and to make it possible to adopt a synchronized scrambling technique.



**FIGURE 1.1** Digital mobile radio communication system.



**FIGURE 1.2** Framed signal.

The next stage of a transmitter is digital modulation. There is no difference in principle between digital modulation and analog modulation—modulation is a process where the amplitude and/or phase of a coherently oscillating carrier signal (although noncoherent carrier as noise is used for some systems) are varied in proportion to a modulating signal. A narrow spectrum bandwidth of the modulated signal is desired. For this purpose, a low-pass filter or an equivalent IF bandpass filter is employed to limit the bandwidth of the modulating signal. Transfer functions, or impulse response of the low-pass filters, are specially designed for digital transmission as discussed in Chapter 2. After the modulation stage, frequency conversion up to a final RF frequency band, power amplification, and transmission via an antenna follows in subsequent stages.

The signal that is received through an antenna is amplified, frequency down-converted, and then band-limited with a bandpass filter. A demodulator outputs a transmitted baseband signal, which is corrupted to some extent by noise and imperfections in the function of the transmitter, channel, and/or receiver. The principle for the demodulation of a digitally modulated signal is the same as for an analog signal, that is, coherent or noncoherent demodulation can be used. In some digital transmission



systems, a carrier signal, which is necessary for coherent demodulation, can be extracted from a received signal by using the knowledge that the digitally modulated signals take specific phases.

The demodulated signal is fed into a decision circuit, where the received signal is sampled and discretized to one of the allowable values of the transmitted signals. A timing recovery circuit generates a timing signal for sampling. The timing signal is extracted from the received signal by detecting the change in digital signal levels at clock frequency duration. The decision error rate is the main concern at the receiving side. Decision is a process particular to a digital transmission system. The principle of the decision process is based on the fact that digital signals take one of several discrete states. If there is no decision error, then noise, interference, or distortion will have no effect on the signal transmission. This fact becomes especially important for a multihop transmission system, where the signal is regenerated and transmitted at the repeaters: no accumulation of noise and interference occurs in the digital system in contrast to an analog system.

Output of the decision circuit takes the form of a logic signal. Channel decoding, descrambling, and deassembling of the framed signal, which are inverse operations to respectively channel coding, scrambling, and framing, are carried out with logic circuits.

The digital signal is converted into an analog voice signal with a voice decoder. As long as a voice signal is involved, the phase of the transfer function is not important, because human ears are not much sensitive to the phase of the voice signal. On the other hand, decision error in the digital transmission is sensitive to both the phase and the amplitude of the transfer function. This is the reason why pulse waveforms should be transmitted without distortion.

The transmission of data signals differs from voice signal transmission in several aspects. We should assure a very low error rate, because great value may depend on accurate data transmission, for example, in a banking service. In contrast to this we can tolerate a rather higher bit error rate for speech communication, because the voice signal has a lot of redundancy, and humans are very intelligent communication terminals: we can ask for repetition and confirm the meaning, if received voice signal is not correct due to transmission error.

The intelligence of the data communication system is given by protocols imbedded in the data terminal. The protocol affects the efficiency of data transmission even in the same transmission channel. In a mobile radio communication channel, in which error rate performance is not generally good, an automatic repeat request with error detection seems indispensable for data transmission. Real-time communication is not particularly important for data transmission, in contrast to spoken conversation.

A mobile radio communication channel is characterized as a channel without direct propagation. When mobile equipment moves quickly, a transmitted signal encounters rapid fading phenomena. The depth of the fading can reach as deep as several tens of dB. The fading speed is proportional to a carrier frequency and the speed of motion. For example, the maximum Doppler frequency is as high as 90 Hz for a carrier frequency of 900 MHz and a car speed of 100 km/h. A receiver has to

cope with this fast fading phenomena. On the other hand, in the case of hand-carried equipment, the fading speed is quite slow because of the low speed of motion. In this case another problem occurs, the fading duration becomes too long: in this case we cannot communicate for a long time. As the digital transmission rate increases, the channel exhibits frequency-selective fading, where the fading is different at different frequencies and, as a result, the transfer function of the channel loses its distortion free characteristic. The distortion in the frequency-selective fading channel degrades bit error performance due to intersymbol interference.

Noise is a general problem in communications. Carrier frequencies for mobile radio communications are predominately in the VHF and UHF bands. At these frequencies, the man-made noise and atmospheric noise are rather high compared with the microwave communication band. For the purpose of obtaining higher receiving sensitivity, it is ineffective to use a low noise receiver which has a noise level lower than the channel noise. Digital transmission offers a solution, that is, an error control scheme, to make communication secure in a noisy channel.

Eavesdropping on a conversation by a third party can occur in mobile radio communications if the transmitted signal is not protected: the transmitted signal propagates everywhere around the transmitter and one can easily get eavesdropping equipment (i.e., a scanning receiver). Voice scramblers for analog transmission are not sufficient to assure the desired degree of protection or voice quality. Digital transmission offers a high security and high voice quality scrambling to combat eavesdropping.

Mobile radio communication equipment is, in some sense, comparable to other consumer products: the number of mobile terminals in the market is large in contrast to other communication systems, such as satellite communication or microwave communication equipment. A mobile terminal is so slim that it slips into pocket and is so inexpensive that consumers can easily afford it. For a digital mobile radio terminal, an increase in the price or size will not be accepted in the market, as long as the service is equivalent.

The biggest difference between mobile radio communications and other consumer products is that the former is supported with a huge system, while the latter is not. Consider a mobile telephone system; for example, the mobile switching centers and a number of base stations, as well as the mobile terminals, are incorporated into the system. Digitization of the mobile telephone system operates on all of these facilities. The switching center and the wire line communications were digitized prior to the digitization of the base stations and the mobile terminals. Communications between the base station and the mobile terminals (i.e., through the air interface) is the major concern of this book. This interface includes digital modulation/demodulation, channel access methods for use of the channels by many users on demand, and other topics related to the signal transmission between the base station and the subscriber stations.

Digital mobile communication service became widespread all over the world. And businesses associating with the digital mobile communications have expanded to contribute to increasing economical activities.

## 1.2 THE PURPOSE OF DIGITIZATION OF MOBILE RADIO COMMUNICATIONS

Data transmission, voice scrambling, and spectrum efficiency are the major motivations for digitization of mobile communications. The emphasis is different for each purpose.

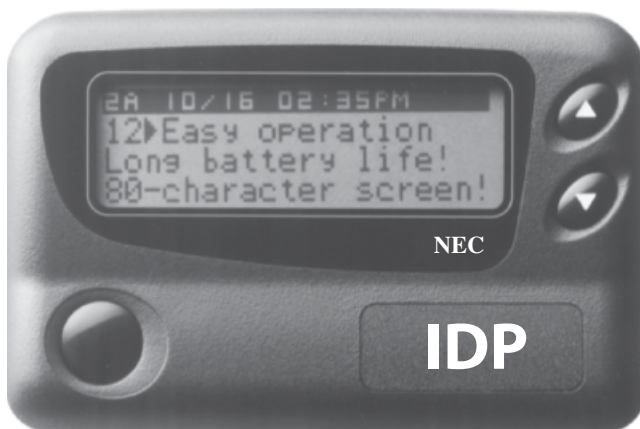
### 1.2.1 Data Communication

Data transmission plays the primary role in controlling a mobile communication system. An example of a controlling system is called set-up and termination control in a mobile telephone system. Although data transmission is relatively slow in speed, high reliability of the transmission is required in this case. Without reliable data transmissions in an unstable mobile radio channel, no advanced mobile communication system could be realized.

In a digital paging system, the radio station broadcasts paging signals in the form of transmitted data. The radio paging system was the earliest example of digitization in mobile communications. The digital radio paging system can accommodate 30,000 subscribers per channel, which is three times larger than conventional tone signaling system.

Computer data transmission between mobile data terminals and host computers has been performed. For example, a sales person may send some business data to the host computer using digital mobile radio communication equipment from his station in his car, and he will be able to have the data automatically processed to yield required items, such as bills or documents, when he returns to the office. For another example, car dispatch service efficiency will be greatly enhanced by adopting a mobile radio data communication system.

Display paging service is a forerunner of mobile message communications (Fig. 1.3). It is expected that message communication through a digital mobile radio channel will be important as electronic mail service becomes more popular. Message



**FIGURE 1.3** Display pager. Courtesy of NEC Corporation.

communication is more advantageous than speech communication from the point of view that it does not disturb the recipient if he or she is engaged in other work and it is more spectrum-efficient than mobile radio speech conversation. The display pager which used to be an advanced system has ceased its role due to the evolution of digital cellular phones.

The integrated services digital network has been put into operation. The application of this network technology will surely be extended in the near future to mobile radio communications. Thus, the digitization of the mobile radio channel is considered imminent in more advanced and versatile communication systems.

The evolution and penetration of the internet greatly changed the telecommunication services. To expand its service into mobile radio communications, new technologies have been introduced. As a measure of performance, the transmission data rate has increased by a million times from 10 kbps at an initial stage to 100 Mbps with latest long term evolution (LTE) system (Section 9.7.6) during last 20 years. The major mobile terminals have changed from voice terminals to so-called smart phones based on the PC functions.

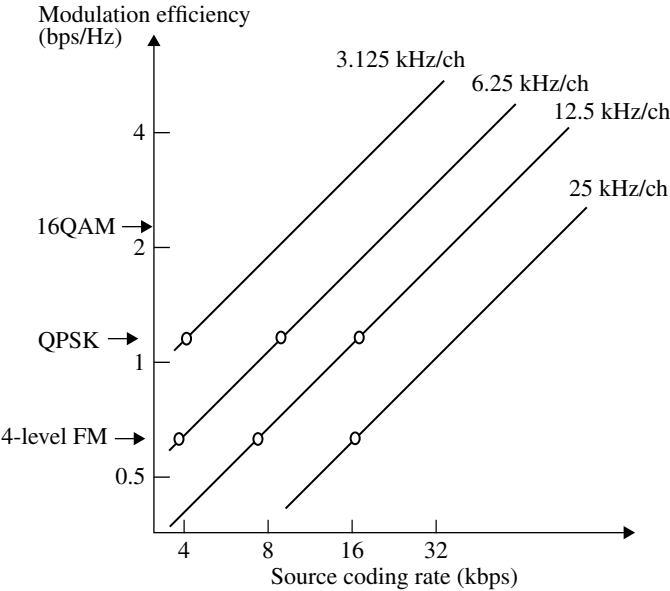
### **1.2.2 Voice Scrambling**

Voice protection from eavesdropping in mobile radio communications is the most important requirement for security in military or police communication. Even in public mobile communications, such as by mobile telephone or cordless telephone, voice scrambling has received increased attention as those services have become widespread. Although some voice protection techniques have been developed for analog communication systems, they do not provide a sufficient degree of protection and/or voice quality. Voice scrambling in digital communication resolves those shortcomings: it is hardly recognizable by a third party whether the signal contains conversation or not, yet nevertheless the voice quality is not degraded.

### **1.2.3 Spectrum Efficiency**

Spectrum efficiency is the most important issue for radio communications, since the radio frequency spectrum band is considered to be a limited natural resource. In general, digital transmission needs a wider bandwidth than analog transmission: if we transmit a voice signal using the 64 kbps pulse code modulation (PCM) in two-level signaling, we need at least 32 kHz of baseband frequency. This is about eight times wider than that needed for analog transmission. The reason why the development of digital technologies for mobile radio communication has been delayed so far, compared with wire line communications and fixed microwave communication, is that the requirement for spectrum efficiency is so tight and that no technology to date has been able to cope with this requirement. Recent progress in digital mobile communication has succeeded in overcoming this problem.

Digital modulation and voice coding play a dominant role in spectrum efficiency. Figure 1.4 shows spectrum efficiency in terms of the required bandwidth per channel or the channel spacing for digital voice transmission as a function of the voice coding



**FIGURE 1.4** Required bandwidth per channel as a function of modulation efficiency and source coding rate.

rate and the efficiency of the digital modulation. The efficiency of digital modulation (bits/s/Hz) is defined as a transmission bit rate (bits/s or bps) per unit spectrum bandwidth (Hz). A 25kHz channel spacing, which is the most conventional, can be achieved with the digital modulation of four-level FM combined with voice coding using 16kbps adaptive delta modulation. If voice coding at a 8kbps coding rate or digital modulation with double the spectral efficiency, such as linear quadrature phase shift keying, is introduced, a 12.5 kHz channel spacing is possible. A 6.25 kHz channel spacing is possible if both the above techniques are utilized. As will be described later, the per-channel bandwidth of 3.125 kHz is possible at this moment.

For a cellular system, there is another factor that governs the spectrum efficiency, that is, frequency- or channel-reuse distance. A shorter reuse distance is desirable. The digital system can reduce the reuse distance by using error correction techniques to mitigate the cochannel interference effect.

The spectral efficiency is enhanced furthermore with MIMO (multi-in multi-output) systems, where multiple data streams can be sent at the same time and frequency by canceling the interchannel interference.

**1.2.4 System Cost**

Cost increases sometimes when a communication system is digitized. However, introducing an *N*-channel time division multiple access scheme for digital transmission, the cost of the transmitter–receiver at the base station may be reduced, because

the number of the transmitters and receivers are decreased to  $1/N$  of the conventional analog system using the frequency division multiple access. Furthermore, by introducing a recently developed common amplifier for TDM systems, the base station's cost has been dramatically reduced. However, the cost of a digital mobile telephone terminal may be higher in that case than that of the conventional analog mobile telephone equipment. But, it is more meaningful to consider the relative terminal cost per spectrum bandwidth, even if the total cost increases. The increase in the total number of channels in the digital system can thus result in much lower cost per spectrum bandwidth.

---

# 2

---

## SIGNAL AND SYSTEMS

This chapter provides a mathematical foundation for the analysis of signals, noise, and linear signal systems that will be used in subsequent chapters. The reader who is familiar with these fundamental techniques may choose to move on to the next chapter.

### 2.1 SIGNAL ANALYSIS

#### 2.1.1 Delta Function

The delta function was invented by a physician, P. A. M. Dirac, to mathematically describe the quantum mechanics [1]. Although the intuitive understanding is rather easy, its mathematical background differs from conventional functions [2]. Therefore, it is called distribution or hyperfunction. The mathematical treatment as hyperfunction is highly abstract and it is hard for engineers (who just apply the delta function to practical matters) to follow its treatment. In some engineering textbooks, the description on the Dirac delta function is not sufficient enough. In this book, we primitively describe the delta function as a (generalized) limit of conventional functions without touching the argument of hyperfunctions. The delta function plays an important role to develop theories of the Fourier analysis and to analyze signal and systems as described in this book.

**2.1.1.1 Definition of the Delta Function** Delta function is defined through integral as follows:

$$\int_{-\infty}^{\infty} f(t)\delta(t)dt = f(0) \quad (2.1)$$

where  $f(t)$  is an arbitrary function being continuous at  $t = 0$ . We consider a following function

$$f(t) = \begin{cases} f_1(t) & (a \leq t \leq b) \\ 0 & (\text{otherwise}) \end{cases}.$$

Inserting this into Equation 2.1, we get the following equation:

$$\int_{-\infty}^{\infty} f(t)\delta(t)dt = \int_a^b f_1(t)\delta(t)dt = \begin{cases} f_1(0) & (ab < 0) \\ 0 & (\text{otherwise}) \end{cases} \quad (2.2)$$

where  $a$  and  $b$  are arbitrary constants that never take zero simultaneously. If  $a$  or  $b$  takes zero, then

$$\int_a^b f_1(t)\delta(t)dt = \frac{f_1(0)}{2}$$

If  $f(t)$  is discontinuous at  $t = 0$ , we have

$$\int_{-\infty}^{\infty} f(t)\delta(t)dt = \frac{\{f(0^+) + f(0^-)\}}{2} \quad (2.3)$$

where  $f(0^+)$  and  $f(0^-)$  are the limits on the right and the left, respectively. If we use the impulse function known as the delta function, that is,  $\delta(t) = 0(t \neq 0)$ , and  $\int_{-\infty}^{\infty} \delta(t)dt = 1$ , then we can show the result for  $a, b \rightarrow 0$  in Equation 2.2. The condition  $\delta(t) = 0(t \neq 0)$  is, however, not necessary as shown in the following.

Following the aforementioned arguments, we may not get sufficient understanding on the delta function. Actually, no conventional function exists to satisfy Equation 2.1. To express the delta function with an ordinary function, we should consider the following (generalized) limit.

$$\lim_{n \rightarrow N} \int_{-\infty}^{\infty} f(t)\delta_n(t)dt = f(0) \quad (2.4)$$

where  $N$  is an extreme and  $\delta_n(t)$  is an ordinary function series.



### 2.1.1.2 Example of Function that Moves to the Delta Function

*The Dirac Type* This is a well-known delta function:

$$\delta_\varepsilon(t) = \begin{cases} \frac{1}{\varepsilon} & \left( |t| \leq \frac{\varepsilon}{2} \right) \\ 0 & (\text{otherwise}) \end{cases}$$

As shown in Figure 2.1, this function becomes  $\delta_\varepsilon(t) = 0$  ( $t \neq 0$ ) and  $\delta_\varepsilon(0) = \infty$  ( $t = 0$ ) for  $\varepsilon \rightarrow 0$  keeping the area size of unity. Since  $\delta_\varepsilon(t)$  ( $\varepsilon \rightarrow 0$ ) never converges at ( $t = 0$ ), the integral  $\int_{-\infty}^{\infty} \lim_{\varepsilon \rightarrow 0} \delta_\varepsilon(t) f(t) dt$  cannot be carried out, if the limiting operation is taken inside of the integral. Moving the limiting operation outside the integral, we get a convergent result as follows:

$$\lim_{\varepsilon \rightarrow 0} \int_{-\infty}^{\infty} f(t) \delta_\varepsilon(t) dt = \lim_{\varepsilon \rightarrow 0} \int_{-\varepsilon/2}^{\varepsilon/2} f(t) \frac{1}{\varepsilon} dt = \lim_{\varepsilon \rightarrow 0} f(0) \int_{-\varepsilon/2}^{\varepsilon/2} \frac{1}{\varepsilon} dt = f(0)$$

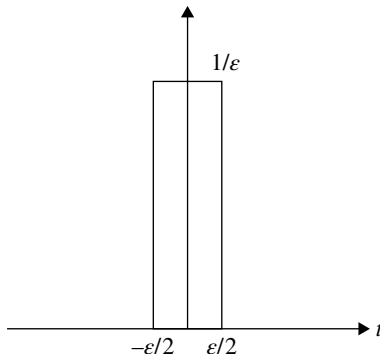
We denote  $\lim_{\varepsilon \rightarrow 0} \delta_\varepsilon(t) = \delta(t)$  in a general sense that the function,  $\lim_{\varepsilon \rightarrow 0} \delta_\varepsilon(t) = \delta(t)$ , satisfies Equation 2.4 where the limit is taken outside the integral. It is easily confirmed that  $\lim_{\varepsilon \rightarrow 0} \delta_\varepsilon(t)$  satisfies Equation 2.3.

*The Dirichlet Integral Type* This is given as follows:

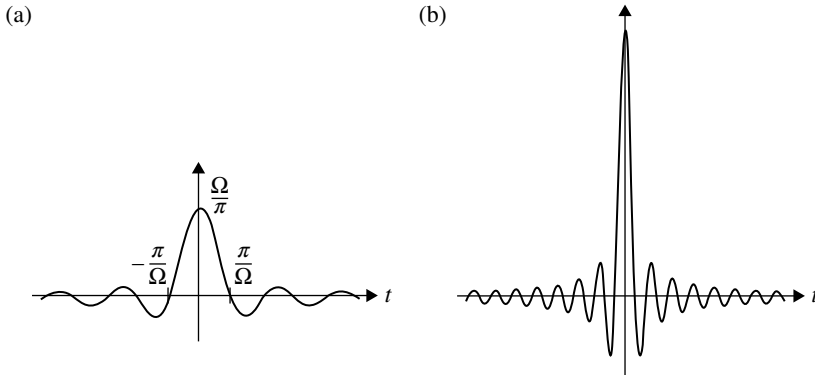
$$\lim_{\Omega \rightarrow \infty} \delta_\Omega(t) = \delta(t), \quad \delta_\Omega(t) = \frac{\sin \Omega t}{\pi t}, \quad (2.5)$$

The integral of  $\delta_\Omega(t)$  becomes unity as shown in the following equation. Using  $x = \Omega t$  we get,

$$\int_{-\infty}^{\infty} \delta_\Omega(t) dt = \int_{-\infty}^{\infty} \frac{\sin x}{\pi x} dx = 1$$



**FIGURE 2.1** An example of function that tends to the delta function.



**FIGURE 2.2** Function,  $\sin \Omega t / \pi t$ . (a)  $\Omega = 2\pi$ , (b)  $\Omega = 6\pi$ .

where we used formula  $\int_{-\infty}^{\infty} \frac{\sin x}{x} dx = \pi$ .

For  $\Omega \rightarrow \infty$ , we have  $\delta_{\Omega}(t) \neq 0$  for almost all  $t$ , and  $\delta_{\Omega}(t)$  never converges for all  $t$  (Fig. 2.2). However, the generalized limit converges and Equation 2.2 yields  $f(0)$  when function  $f(t)$  fulfills conditions given in Appendix 2.A.

It is shown in Appendix 2.A that Equation 2.3 is valid also for  $\lim_{\Omega \rightarrow \infty} \delta_{\Omega}(t)$ .  $\delta_{\Omega}(t)$  can be given as follows:

$$\delta_{\Omega}(t) = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} e^{j\omega t} d\omega = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \cos \omega t d\omega$$

Thus,  $\lim_{\Omega \rightarrow \infty} \delta_{\Omega}(t)$  is given by adding (integrating) trigonometric functions of all frequencies in phase.

We have many functions that tend to the delta function, for example, as follows:

$$\lim_{\sigma \rightarrow 0} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} (-\infty < x < \infty)$$

**2.1.1.3 The Properties of the Delta Function** If we assume  $t$  as a probabilistic variable, examples described in Section 2.1.1.2 are considered as a probability density function,  $p(t)$ . Then, probability for  $t_1 \leq t \leq t_2$  becomes  $\int_{t_1}^{t_2} p(t) dt$  and all the probabilities become  $\int_{-\infty}^{\infty} p(t) dt = 1$  ( $f(t) = 1$  in Eq. 2.1). Thus, Equation 2.1 can be considered as a formula to calculate a probability average. The probabilities given by the Dirac type delta function is concentrated at one point of  $t=0$ . The probabilities given by the Dirichlet type delta function are distributed in infinite region. However, the average probability given with the Dirichlet type delta function becomes the same as that of the

Dirac type delta function, where the probabilities are concentrated at one point. From this sense also, the delta function  $\delta(t)$  is called distribution or hyperfunction differentiating from ordinary functions. Although  $\delta(t)$  given with examples as the Dirac type and the Dirichlet type behaves quite differently, the calculated results with Equation 2.4 (generalized extreme) become the same.

### 2.1.1.4 Formulas on the Delta Function

$$\int_{-\infty}^{\infty} \delta(t) dt = 1 \quad (2.6)$$

$$\int_{-\infty}^{\infty} f(t) \delta(t - t_0) dt = f(t_0) \quad (2.7)$$

$$\int_{-\infty}^{\infty} \delta(at) f(t) dt = \frac{1}{|a|} f(0) \quad (2.8)$$

From Equation 2.8, we have  $\delta(-t) = \delta(t)$ . Thus,  $\delta(t)$  is an even function. With this property and Equation 2.6, we have  $\int_{-\infty}^0 \delta(t) dt = \int_0^{\infty} \delta(t) dt = \frac{1}{2}$ . In order for the half-infinite integral to converge, we should assume  $\delta(\pm\infty) = 0$ .

$$f(t) \delta(t) = f(0) \delta(t) \quad (2.9)$$

We understand the above equation means  $\int_{-\infty}^{\infty} f(t) \delta(t) dt = \int_{-\infty}^{\infty} f(0) \delta(t) dt (= f(0))$ . Since the delta function has physical meanings after integration, we use a formal expression as Equation 2.9.

$$\int_{-\infty}^{\infty} f(t) \frac{d}{dt} \delta(t) dt = -f'(0)$$

This is given by setting the first term to be zero in the following integral by part.

$$\int_{-\infty}^{\infty} f(t) \frac{d}{dt} \delta(t) dt = [f(t) \delta(t)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f'(t) \delta(t) dt$$

For this purpose, it is required that  $f(\pm\infty) < \infty$  ( $\delta(\pm\infty) = 0$ ). Similarly, for  $n$ th differentiation, we get

$$\int_{-\infty}^{\infty} f(t) \frac{d^n}{dt^n} \delta(t) dt = (-1)^n f^{(n)}(0)$$

Although  $\delta^2(t)$  is not defined, the convolutional integral (definition: Eq. 2.38) becomes as:

$$\delta(t - t_1) * \delta(t - t_2) = \int_{-\infty}^{\infty} (\tau - t_1) \delta(\tau - t_1 - t_2) d\tau = \delta(t - t_1 - t_2)$$

**2.1.1.5 Integral Representation of Functions with the Delta Function** A function  $f(t)$  is step-approximated as follows:

$$f(t) \cong \sum_{n=-\infty}^{\infty} f(n\Delta T) p_{\Delta T}(t - n\Delta T)$$

where,

$$p_{\Delta T}(t) = \begin{cases} 1 & |t| \leq \frac{\Delta T}{2} \\ 0 & \text{otherwise} \end{cases}.$$

From the above equation, by denoting  $\Delta T = d\tau$ ,  $n\Delta T = \tau$  for  $\Delta T \rightarrow 0$ , we have the following convolutional integral:

$$\begin{aligned} & \lim_{\Delta T \rightarrow 0} \sum_{n=-\infty}^{\infty} f(n\Delta T) p_{\Delta T}(t - n\Delta T) \Delta T / \Delta T \\ &= \int_{-\infty}^{\infty} f(\tau) \delta(t - \tau) d\tau = f(t) \end{aligned}$$

where  $\lim_{\Delta T \rightarrow 0} p_{\Delta T}(t)/\Delta T = \delta(t)$  is used.

**2.1.1.6 Representation of Differentiation at a Discontinuous Point with the Delta Function** Consider a function  $f(t)$  that is discontinuous at  $t = t_0$  with a jump of  $J_0$ . Then, we have the following relation:

$$f(t) = f_c(t) + J_0 u(t - t_0) \quad (2.10)$$

where  $f_c(t)$  is a continuous function and  $u(t)$  is the unit step function (Eq. 2.46). We denote the differentiation by  $f'(t)$  and then

$$\int_{t_0 - \varepsilon}^{t_0 + \varepsilon} f'(t) dt = J_0 \quad (\varepsilon \rightarrow 0).$$

Thus,  $f'(t)$  is represented with the delta function as follows:

$$f'(t) = f'_c(t) + J_0 \delta(t - t_0)$$

With this equation and Equation 2.10, we have the following relation:

$$\frac{d}{dt} u(t) = \delta(t)$$

Integrating this equation and using Equation 2.2, the following relation is given:

$$u(t) = \int_{-\infty}^t \delta(\tau) d\tau$$

Considering  $\int_{-\infty}^0 \delta(\tau) d\tau = 1/2$ , we have  $u(0) = 1/2$ .

## 2.1.2 Fourier Analysis

Here, a short explanation of the Fourier series and the Fourier integral is given. The Fourier integral is treated as an expansion of definition region of the Fourier series. Then, the Fourier series are discussed again as the Fourier integral of a periodic function.

**2.1.2.1 The Fourier Series** We consider a function  $f(t)$  defined in a region  $(-T/2 \sim T/2)$ . The Fourier series is given as:

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t} \quad \left( -\frac{T}{2} \leq t \leq \frac{T}{2} \right) \quad (2.11)$$

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-jn\omega_0 t} dt \quad (2.12)$$

where,  $\omega_0 = 2\pi/T$ . Coefficient  $c_n$  is called *spectrum*.

We show that the coefficient  $c_n$  is given as in Equation 2.12, in the following. Multiplying  $(1/T)e^{-jm\omega_0 t}$  with both sides of Equation 2.11, and integrating in a region  $-T/2 \sim T/2$ , we have:

$$\frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-jm\omega_0 t} dt = \sum_{n=0}^{\infty} \frac{c_n}{T} \int_{-T/2}^{T/2} e^{j(n-m)\omega_0 t} dt = c_m \quad \left( \omega_0 = \frac{2\pi}{T} \right)$$

where the following orthogonal relation is used:

$$\int_{-T/2}^{T/2} e^{j(n-m)\omega_0 t} dt = \begin{cases} T & m = n \\ 0 & m \neq n \end{cases}$$

Thus, the Fourier series is a orthogonal function expansion with trigonometric functions. Letting  $c_n = (a_n - jb_n)/2$  for real  $a_n, b_n$ , and  $c_{-n} = c_n^*$ , Equation 2.11 can be rewritten as follows:

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos n\omega_0 t + b_n \sin n\omega_0 t]$$

Now we examine if the series expansion converges uniquely to  $f(t)$ . To this end, we insert Equation 2.12 to the right side of Equation 2.11 and show that

$$\lim_{N \rightarrow \infty} \sum_{n=-N}^N \frac{1}{T} \int_{-T/2}^{T/2} f(x) e^{-jn\omega_0 x} dx e^{jn\omega_0 t} \rightarrow f(t).$$

Replacing the order of the integral and summation, we have the following equation.

$$\lim_{N \rightarrow \infty} \sum_{n=-N}^N \frac{1}{T} \int_{-T/2}^{T/2} f(x) e^{jn\omega_0(t-x)} dx = \lim_{N \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(x) D_N(t-x) dx \quad (2.13)$$

where  $D_N(t) = \sum_{n=-N}^N e^{jn\omega_0 t}$  is called the Dirichlet kernel.  $D_N(t)$  is a periodic function of period  $T$  and is rewritten as follows:

$$D_N(t) = \frac{e^{-jN\omega_0 t} - e^{j(N+1)\omega_0 t}}{1 - e^{j\omega_0 t}} = \frac{\sin\{(N+1/2)\omega_0 t\}}{\sin(\omega_0 t/2)}$$

The term  $D_N(t)/T$  is further modified as follows:

$$\frac{D_N(t)}{T} = \frac{1}{T} \frac{\pi t}{\sin(\omega_0 t/2)} \frac{\sin \Omega t}{\pi t} = \frac{\omega_0 t/2}{\sin(\omega_0 t/2)} \frac{\sin \Omega t}{\pi t}$$

where we put  $\Omega = (N+1/2)\omega_0$ . By the way, we know  $\lim_{N \rightarrow \infty} \sin \Omega t / \pi t = \delta(t)$  (Appendix 2.A). Since  $\frac{\omega_0 t/2}{\sin(\omega_0 t/2)}$  satisfies the conditions for a test function for the delta function given by  $\lim_{\Omega \rightarrow \infty} \sin \Omega t / \pi t$  in a period of  $|t| < T/2$ , we have the following relation:

$$\lim_{N \rightarrow \infty} \frac{1}{T} D_N(t) = \frac{\omega_0 t/2}{\sin(\omega_0 t/2)} \delta(t) = \delta(t) \quad \left( |t| < \frac{T}{2} \right)$$

where a property of the delta function  $f(t)\delta(t) = f(0)\delta(t)$  (Eq. 2.9) is used. (For  $-\infty \leq t \leq \infty$ , we have  $\lim_{N \rightarrow \infty} (1/T) D_N(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$ .) Applying this result to Equation 2.13, we get the following relation:

$$\lim_{N \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(x) D_N(t-x) dx = \int_{-T/2}^{T/2} f(x) \delta(t-x) dx = f(t)$$

where we used the properties of the delta function given with Equations 2.2 and 2.7. We should recall that  $f(t)$  should fulfill the conditions (Appendix 2.B) as a test function to define the delta function given by  $\lim_{\Omega \rightarrow \infty} \sin \Omega t / \pi t$ .

At a discontinuous point,  $f(t)$  converges to an average value of right- and left-side extremes  $\{f(t^+) + f(t^-)\} / 2$  (Eq. 2.3).

**2.1.2.2 The Fourier Integrals** Introducing a variable  $F(n\omega_0) = Tc_n$ , Equations 2.11 and 2.12 can be modified as follows:

$$f(t) = \sum_{n=-\infty}^{\infty} \frac{1}{T} F(n\omega_0) e^{jn\omega_0 t} \quad \left( -\frac{T}{2} \leq t \leq \frac{T}{2} \right)$$

$$F(n\omega_0) = \int_{-T/2}^{T/2} f(t) e^{-jn\omega_0 t} dt$$

where  $\omega_0 = 2\pi / T$ . For  $T \rightarrow \infty$  letting  $2\pi / T = d\omega$  and  $n\omega_0 = \omega$ , the above equations move to integrals as follows:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega \quad (-\infty \leq t \leq \infty) \quad (\text{the inverse Fourier transform}) \quad (2.14)$$

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (\text{the Fourier transform}) \quad (2.15)$$

Relation between  $f(t)$  and  $F(\omega)$  is called the Fourier transform pair and is expressed later as  $f(t) \leftrightarrow F(\omega)$ .

For the Fourier transform to exist, the above two equations should hold simultaneously. We will show that insertion of Equation 2.15 into Equation 2.14 leads to  $f(t)$ . For the sufficient condition, it is known that  $f(t)$  is absolutely integrable, that is,  $\int_{-\infty}^{\infty} |f(t)| dt < \infty$  [2]. This condition is too strict for a certain function, for example, a trigonometric function,  $f(t) = A \cos \omega_0 t$  ( $-\infty \leq t \leq \infty$ ), which is important in engineering field. To avoid the condition, we introduce restrictions when moving to infinity in the Fourier transform as follows:

$$f(t) = \lim_{T, \Omega \rightarrow \infty} \frac{1}{2\pi} \int_{-\Omega}^{\Omega} F(\omega) e^{j\omega t} d\omega \quad (2.16)$$

$$F(\omega) = \lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} f(t) e^{-j\omega t} dt \quad (2.17)$$

In the conventional definition of the Fourier integral (Eqs. 2.14 and 2.15), the upper and lower limits are taken independently. In our definition (Eqs. 2.16 and 2.17), we take the limits symmetrically (the Cauchy principal value integration). Furthermore, we introduce the other restriction as  $\lim_{T, \Omega \rightarrow \infty} T/\Omega = 0$ , taking into

consideration the fact that  $n\omega_0$  becomes infinite for a finite  $T$  in the Fourier series expansion. When applying the Fourier integrals to engineering subjects, our restrictions never cause any inconvenience. Inserting Equation 2.17 into Equation 2.16 and replacing the order of integration, we have the following relation:

$$\begin{aligned} \lim_{T, \Omega \rightarrow \infty} \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \int_{-T/2}^{T/2} f(x) e^{-j\omega x} dx e^{j\omega t} d\omega &= \lim_{T, \Omega \rightarrow \infty} \int_{-T/2}^{T/2} f(x) dx \frac{1}{2\pi} \int_{-\Omega}^{\Omega} e^{j\omega(t-x)} d\omega \\ &= \lim_{T, \Omega \rightarrow \infty} \int_{-T/2}^{T/2} f(x) \frac{\sin \Omega(t-x)}{\pi(t-x)} dx \end{aligned}$$

Since  $\lim_{\Omega \rightarrow \infty} \frac{\sin \Omega(t-x)}{\pi(t-x)} = \delta(t-x)$  (Eq. 2.5), the above equation becomes

$[f(t^+) + f(t^-)]/2$ . Here, the conditions required for  $f(t)$  are that  $f(t)$  and  $f'(t)$  are, for example, bounded as shown in Appendices 2.A and 2.B. Thus, the Fourier integrals under our restrictions are applicable to functions that are not absolutely integrable as  $f(t) = e^{j\omega t}$  ( $-\infty \leq t \leq \infty$ ).

If we use  $f = \omega/2\pi$  instead of  $\omega$ , we have a symmetric expression without the coefficient  $1/2\pi$ , as follows:

$$\begin{aligned} f(t) &= \int_{-\infty}^{\infty} F(2\pi f) e^{j2\pi f t} df \\ f(2\pi f) &= \int_{-\infty}^{\infty} f(t) e^{-j2\pi f t} dt \end{aligned}$$

The spectrum  $c_n$  becomes continuous for  $T \rightarrow \infty$ . We denote this as  $c(\omega)$ , then we have:

$$F(\omega) = \frac{2\pi c(\omega)}{d\omega} = \frac{c(2\pi f)}{df} \quad (\omega = 2\pi f)$$

This relation means that  $F(\omega)$  is the density of  $c(\omega)$ , that is, the spectral density. If the unit of  $c(\omega)$  is voltage and that of  $\omega$  is the angular frequency, the unit of  $F(\omega)$  becomes volt per Hz (V/Hz).

It happens that the spectral density  $F(\omega)$  may become infinite and its integral may stay to be finite. For an example, consider a function  $f(t) = e^{j\omega_1 t}$  ( $-\infty \leq t \leq \infty$ ), then we have  $F(\omega) = 2\pi\delta(\omega - \omega_1)$  (Eq. 2.22). The spectral density becomes infinite at  $\omega = \omega_1$ ; however, the coefficient,  $c_n \equiv c(\omega)$ , of the Fourier series is constant for  $T \rightarrow \infty$  and is as follows:

$$c_n = \begin{cases} 1 & n\omega_0 = \omega_1 \\ 0 & n\omega_0 \neq \omega_1 \end{cases} \quad (T \rightarrow \infty)$$



And the Fourier series converge uniquely. The singularity that the infinity appears in the Fourier integral is caused by the fact that we introduced the density function,  $F(n\omega_0) = Tc_n$ , in moving from the Fourier series expansion to the Fourier integral.

**2.1.2.3 The Fourier Integrals of Periodic Functions** A periodic function of a period of  $T$  is expressed as follows:

$$f(t) = \sum_{n=-\infty}^{\infty} f_T(t-nT) = f_T(t) * \sum_{n=-\infty}^{\infty} \delta(t-nT)$$

where symbol “\*” means the convolutional integral (defined in Eq. 2.37) and  $f_T(t) = 0$  for  $|t| > T/2$  is assumed. A periodic function of the delta function  $\sum_n \delta(t-nT)$  is called the sampling function. To prepare for the following discussion, we first show that the Fourier transform of the sampling function also becomes a periodic function (Fig. 2.3) of the delta function in frequency domain as follows:

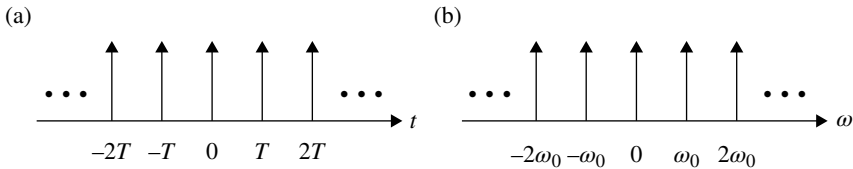
$$\sum_{n=-\infty}^{\infty} \delta(t-nT) \leftrightarrow \omega_0 \sum_{n=-\infty}^{\infty} \delta(\omega-n\omega_0) \quad \left( \omega_0 = \frac{2\pi}{T} \right) \quad (2.18)$$

Proof: The Fourier Transform of  $\sum_{n=-\infty}^{\infty} \delta(t-nT)$  becomes

$$\sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(t-nT) e^{-j\omega t} dt = \sum_{n=-\infty}^{\infty} e^{-jn\omega T}$$

This function is a periodic function of a period  $\omega_0 = 2\pi/T$ . Therefore, we will show that this function becomes  $\omega_0 \delta(\omega)$  for  $|\omega| \leq \omega_0/2$ . We modify  $\sum_{n=-N}^N e^{-jn\omega T}$  as follows:

$$\sum_{n=-N}^N e^{-jn\omega T} = \frac{e^{-jN\omega T} - e^{-j(N+1)\omega T}}{1 - e^{-j\omega T}} = \frac{\sin\{(N+1/2)\omega T\}}{\sin(\omega T/2)} = \frac{\omega T/2}{\sin(\omega T/2)} \frac{2\pi \sin\{(N+1/2)\omega T\}}{\pi \omega T}$$



The delta function series  $\sum \delta(t-nT)$

The Fourier transform  $\omega_0 \sum \delta(\omega-n\omega_0)$

**FIGURE 2.3** The Fourier transform of a periodic delta function ( $\omega_0 = 2\pi/T$ ).

Since  $\lim_{N \rightarrow \infty} \frac{\sin\{(N+1/2)\omega T\}}{\pi\omega T} = \delta(\omega T) = \frac{\delta(\omega)}{T}$  (Eqs. 2.5 and 2.8), we get:

$$\sum_{n=-\infty}^{\infty} e^{-jn\omega T} = \lim_{N \rightarrow \infty} \sum_{n=-N}^N e^{-jn\omega T} = \frac{\omega T/2}{\sin(\omega T/2)} 2\pi \frac{\delta(\omega)}{T} = \omega_0 \delta(\omega) \quad \left( |\omega| \leq \frac{\omega_0}{2} \right) \quad (2.19)$$

where the property of the delta function (Eq. 2.9) is used. The above proof becomes simpler if we use the result of the Fourier transform (series) (Eqs. 2.20 and 2.21), which is shown later.

Now we return to get the Fourier transform,  $F(\omega) \leftrightarrow f(t)$ . From the above result, properties of the Fourier transform (Eq. 2.39) and the delta function (Eq. 2.9), we have:

$$F(\omega) = F_T(\omega) \cdot \omega_0 \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0) = \frac{2\pi}{T} \sum_{n=-\infty}^{\infty} F_T(n\omega_0) \delta(\omega - n\omega_0)$$

where  $F_T(\omega) \leftrightarrow f_T(t)$ . Taking the inverse Fourier transform of the above equation, we get the following equation:

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t} \quad (-\infty \leq t \leq \infty) \quad (2.20)$$

where,

$$c_n = \frac{1}{T} F_T(n\omega_0) = \frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-jn\omega_0 t} dt \quad (2.21)$$

This result is the same as the Fourier series expansion (Eqs. 2.11 and 2.12) except that the time is defined in an infinite region.

### EXAMPLE 2.1

$$e^{j\omega_0 t} \leftrightarrow 2\pi \delta(\omega - \omega_0) \quad (2.22)$$

*Proof*

$$\begin{aligned} F(\omega) &= \lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} e^{j\omega_0 t} e^{-j\omega t} dt \\ &= \lim_{T \rightarrow \infty} \frac{2 \sin(\omega - \omega_0)T/2}{\omega - \omega_0} = 2\pi \delta(\omega - \omega_0) \quad (\text{Eq. 2.5}) \end{aligned}$$

This is the delta function of the Dirichlet integral type.

If we assume a further restriction that  $\omega = n\omega_0$  and  $\omega_1 = m\omega_0$  ( $\omega_0 = 2\pi/T$ ) in moving from the Fourier series to the Fourier transform, then from the orthogonality of the trigonometric functions, we have:

$$F(\omega) = \lim_{T \rightarrow \infty} \begin{cases} T & (\omega = \omega_1) \\ 0 & (\omega \neq \omega_1) \end{cases}$$

This is the delta function of the Dirac type. If we do not take the principal value integral,  $F(\omega)$  is never fixed. Taking the inverse Fourier transform, we confirm the following equation:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} 2\pi\delta(\omega - \omega_0) e^{j\omega t} d\omega = e^{j\omega_0 t}$$

Applying  $\cos \omega t = (e^{j\omega t} + e^{-j\omega t})/2$  and  $\sin \omega t = (e^{j\omega t} - e^{-j\omega t})/2j$  to Equation 2.22, we have the following relation:

$$\cos \omega_0 t \leftrightarrow \pi[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)] \quad (2.23)$$

$$\sin \omega_0 t \leftrightarrow -j\pi[\delta(\omega - \omega_0) - \delta(\omega + \omega_0)] \quad (2.24)$$

Putting  $\omega_0 = 0$  in Equation 2.22, We get:

$$1 \leftrightarrow 2\pi\delta(\omega) \quad (2.25)$$

From the symmetry property of the Fourier transform or the definition of the delta function, we have:

$$\delta(t) \leftrightarrow 1$$

#### 2.1.2.4 The Properties of the Fourier Transform

(i) Linearity

$$a_1 f_1(t) + a_2 f_2(t) \leftrightarrow a_1 F_1(\omega) + a_2 F_2(\omega) \quad (2.26)$$

(ii) Symmetry

$$F(t) \leftrightarrow 2\pi f(-\omega) \quad (2.27)$$

(iii) Compression or expansion of time

$$f(at) \leftrightarrow \frac{1}{|a|} F\left(\frac{\omega}{a}\right) \quad (2.28)$$

(iv) Time shift

$$f(t - t_0) \leftrightarrow F(\omega) e^{-j\omega t_0} \quad (2.29)$$

(v) Frequency shift

$$f(t)e^{j\omega_0 t} \leftrightarrow F(\omega - \omega_0) \quad (2.30)$$

From the above result, we have the following relations for amplitude modulation (AM) signals:

$$f(t)\cos\omega_0 t \leftrightarrow \frac{F(\omega - \omega_0) + F(\omega + \omega_0)}{2} \quad (2.31)$$

$$f(t)\sin\omega_0 t \leftrightarrow \frac{F(\omega - \omega_0) - F(\omega + \omega_0)}{2j} \quad (2.32)$$

(vi) Time differentiation

$$\frac{d^n f(t)}{dt^n} \leftrightarrow (j\omega)^n F(\omega) \quad (2.33)$$

(vii) Time integral

$$g(t) = \int_{-\infty}^t f(\tau) d\tau \leftrightarrow F(0)\pi\delta(\omega) + \frac{F(\omega)}{j\omega} \quad (2.34)$$

(viii) Frequency differentiation

$$(-jt)^n f(t) \leftrightarrow \frac{d^n F(\omega)}{d\omega^n} \quad (2.35)$$

(ix) Complex conjugate function

$$f^*(t) \leftrightarrow F^*(-\omega) \quad (2.36)$$

### 2.1.2.5 The Fourier Transform for Special Cases

(i) *Real Function*

For a real function  $f(t)$ , we have

$$F(-\omega) = F^*(\omega) \quad (2.37)$$

(ii) *The Fourier Transform of the Convolution Integral*

The convolution integral of functions  $f_1(t)$  and  $f_2(t)$  is defined as follows:

$$f_1(t) * f_2(t) = \int_{-\infty}^{\infty} f_1(t-x)f_2(x)dx = \int_{-\infty}^{\infty} f_1(x)f_2(t-x)dx \quad (2.38)$$

Then we can show that

$$f_1(t) * f_2(t) \leftrightarrow F_1(\omega)F_2(\omega) \quad (2.39)$$

(iii) *The Fourier Integral of a Product of Functions*

$$f_1(t)f_2(t) \leftrightarrow \frac{1}{2\pi} F_1(\omega) * F_2(\omega) \quad (2.40)$$

$$F_1(t)F_2(t) \leftrightarrow 2\pi f_1(-\omega) * f_2(-\omega) \quad (2.41)$$

(iv) *The Parseval's Equation*

We rewrite Equation 2.40 as follows:

$$\int_{-\infty}^{\infty} f_1(t)f_2(t)e^{-j\omega t} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_1(\omega - x)F_2(x) dx$$

Letting  $\omega = 0$  in the above equation, we have the following equation known as the Parseval's formula.

$$\int_{-\infty}^{\infty} f_1(t)f_2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_1(-x)F_2(x) dx \quad (2.42)$$

Especially for  $f_2^*(t) = f_1(t)$ , we get the following relation:

$$\begin{aligned} \int_{-\infty}^{\infty} |f(t)|^2 dt &= \frac{1}{2\pi} \int_{-\infty}^{\infty} F^*(\omega)F(\omega) d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega \end{aligned} \quad (2.43)$$

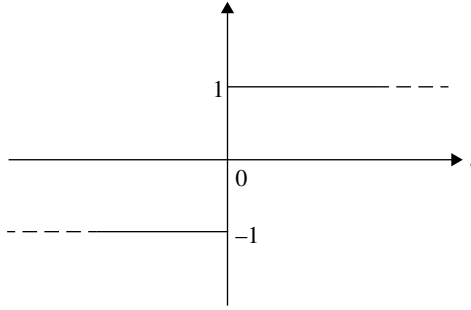
where  $f^*(t) \leftrightarrow F^*(-\omega)$  (Eq. 2.36) is used.

### 2.1.2.6 Examples of the Fourier Transform

$$(1) \quad \text{sgn}(t) \leftrightarrow \frac{2}{j\omega} \quad (2.44)$$

This function takes 1 for  $t > 0$  and  $-1$  for  $t < 0$  (Fig. 2.4), that is,

$$\text{sgn}(t) = \frac{t}{|t|} = \begin{cases} 1 & t > 0 \\ -1 & t < 0 \end{cases}$$



**FIGURE 2.4** Function  $\text{sgn}(t)$ .

Taking the Fourier transform, we have:

$$\begin{aligned} F(\omega) &= \lim_{T \rightarrow \infty} \int_{-T}^T \text{sgn}(t) e^{-j\omega t} dt = \lim_{T \rightarrow \infty} \left[ \frac{e^{-j\omega T} - 1}{-j\omega} - \frac{1 - e^{j\omega T}}{-j\omega} \right] \\ &= \lim_{T \rightarrow \infty} \frac{2 - e^{-j\omega T} - e^{j\omega T}}{j\omega} = \lim_{T \rightarrow \infty} \frac{2 - 2 \cos \omega T}{j\omega} \end{aligned}$$

Since  $\lim_{T \rightarrow \infty} \cos \omega T = 0$  (Eq. 2.B.3), We get the required result.

Let us make the inverse Fourier transform:

$$f(t) = \lim_{\Omega \rightarrow \infty} \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \frac{2}{j\omega} e^{j\omega t} d\omega = \frac{2}{\pi} \int_0^{\infty} \frac{\sin \omega t}{\omega} d\omega \quad (2.45)$$

From the following mathematical formula,

$$\int_{-\infty}^0 \frac{\sin ax}{x} dx = \int_0^{\infty} \frac{\sin ax}{x} dx = \begin{cases} \pi/2 & (a > 0) \\ -\pi/2 & (a < 0) \end{cases}$$

we have the required result as follows:

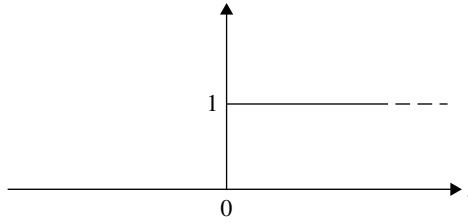
$$f(t) = \begin{cases} 1 & t > 0 \\ -1 & t < 0 \end{cases}$$

If we let  $t=0$  in Equation 2.45,  $f(0)=0$  is obtained.

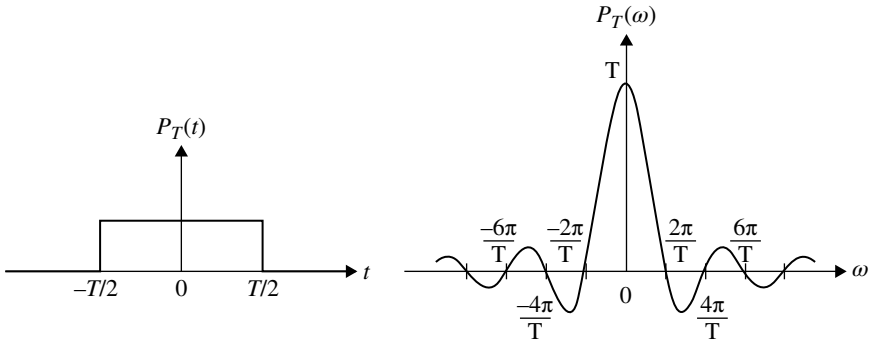
$$(2) \quad u(t) \leftrightarrow \pi \delta(\omega) + \frac{1}{j\omega} \quad (2.46)$$

The step function  $u(t)$  (Fig. 2.5) is defined as follows:

$$u(t) = \frac{1}{2} + \frac{1}{2} \text{sgn}(t) = \begin{cases} 1 & t > 0 \\ 0 & t < 0 \end{cases}$$



**FIGURE 2.5** The step function  $u(t)$ .



**FIGURE 2.6** The Fourier transform of  $p_T(t)$ .

From Equations 2.44 and 2.25, the required result is given. We have the same result directly as:

$$\int_0^{\infty} e^{-j\omega t} dt = \lim_{T \rightarrow \infty} \frac{e^{-j\omega T} - 1}{-j\omega} = \pi\delta(\omega) + \frac{1}{j\omega} \left( \lim_{T \rightarrow \infty} \frac{e^{-j\omega T}}{-j\omega} = \pi\delta(\omega) \right)$$

The inverse Fourier transform can be verified similarly to the case of case (1).

$$(3) \quad p_T(t) \leftrightarrow \frac{T \sin \omega T/2}{\omega T/2} \quad (2.47)$$

where,  $p_T(t) = \begin{cases} 1 & |t| \leq T/2 \\ 0 & \text{otherwise} \end{cases}$ . Although the spectrum decays, it spreads into an infinite region (Fig. 2.6).

For a given function, we have  $f(t)p_T(t) = 0$  ( $|t| \geq T/2$ ). From the property of the Fourier transform of a product of functions (Eq. 2.40), we have the following relation:

$$f(t)p_T(t) \leftrightarrow \frac{1}{2\pi} F(\omega) * \frac{2 \sin \omega T/2}{\omega}$$

From the property of the convolution integral, the spread of the spectrum given in the above expression becomes the sum of the spreads of the two spectra. Since the spectrum spread of  $\sin(\omega T/2)/\omega$  is infinite, the spectrum,  $F(\omega) * \sin(\omega T/2)/\omega$ , spreads into an infinite region even though the spectrum spread of  $F(\omega)$  is finite. Therefore, a function,  $f(t)$ , which becomes zero outside a finite time region, shows an infinite spectrum spread. On the other hand, a spectrum, which is zero outside a finite frequency region, has an infinite spread in time domain.

$$(4) \quad \text{Gauss functions} \quad e^{-a^2 t^2} \leftrightarrow \frac{\sqrt{\pi}}{a} e^{-\left(\frac{\omega}{2a}\right)^2} \quad (2.48)$$

The Fourier transform of a Gaussian function is also a Gaussian function [2].

### 2.1.3 Signals

Here, we first discuss the energy, power, and their spectral densities. Then the correlation function and orthogonal signal are described.

**2.1.3.1 Energy and Power of Signal** The energy of a real signal  $f(t)$  is given as follows:

$$E = \int_{-\infty}^{\infty} f^2(t) dt$$

The energy of an everlasting signal becomes infinite. Let us denote a function,  $f_T(t)$ , which is made by cutting out the everlasting signal  $f(t)$  in a time period,  $(-T \sim T)$ , that is,

$$f_T(t) = \begin{cases} f(t) & -T \leq t \leq T \\ 0 & \text{otherwise} \end{cases}$$

Then, the signal average power is expressed as follows:

$$P = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-\infty}^{\infty} f_T^2(t) dt = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f^2(t) dt$$

**2.1.3.2 The Energy Spectral Density and the Power Spectrum Density** Consider a real signal  $f(t)$  and  $f(t) \leftrightarrow F(\omega)$ . We showed in Section 2.1.2 that  $F(\omega)$  shows spectral density. From the Parseval's equation (Eq. 2.42), we have:

$$E = \int_{-\infty}^{\infty} f^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega \quad (2.49)$$



From the above result, we understand that  $|F(\omega)|^2$  expresses the energy spectral density. Similarly, since

$$P = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-\infty}^{\infty} f_T^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{|F_T(\omega)|^2}{2T} d\omega \quad (f_T(t) \leftrightarrow F_T(\omega)) \quad (2.50)$$

the power spectral density (PSD)  $S(\omega)$  is given as follows:

$$S(\omega) = \lim_{T \rightarrow \infty} \frac{|F_T(\omega)|^2}{2T} \quad (2.51)$$

An equipment that measures the PSD is known as a spectrum analyzer.

**2.1.3.3 The Cross Correlation Function** The (time) correlation function of real functions  $s_i(t)$  and  $s_j(t)$  is defined as follows:

$$R_{ij}(\tau) = \int_{-\infty}^{\infty} s_i(t) s_j(t + \tau) dt$$

This expression can be rewritten as:

$$R_{ij}(\tau) = \int_{-\infty}^{\infty} s_i(-x) s_j(\tau - x) dx = s_i(-\tau) * s_j(\tau)$$

where the symbol “\*” means the convolution integral and its definitions are given in Equation 2.38. For everlasting  $s_i(t)$  and  $s_j(t)$ , the above correlation may take an infinite value. In such a case, we use a definition as follows:

$$R_{ij}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T s_i(t) s_j(t + \tau) dt$$

**2.1.3.4 The Autocorrelation Function** The autocorrelation function of a signal  $s(t)$  is defined as follows:

$$\begin{aligned} R(\tau) &= \int_{-\infty}^{\infty} s(t) s(t + \tau) dt \\ &= \int_{-\infty}^{\infty} s(-x) s(\tau - x) dx = s(-\tau) * s(\tau) \end{aligned}$$

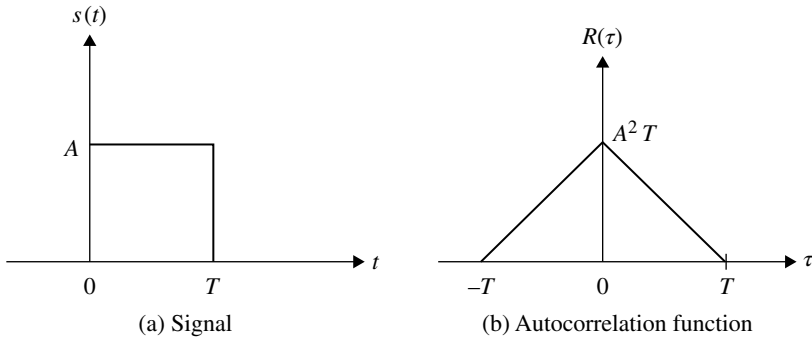
From the above equation, we know  $R(-\tau) = R(\tau)$ . If  $R(\tau) = \delta(\tau)$ , then the energy spectral densities are constant (flat) over  $\omega$ , as given by Equation 2.52 and  $\delta(t) \leftrightarrow 1$ .

For an everlasting signal  $s(t)$ , the autocorrelation function  $R(\tau)$  becomes infinite. In such a case, we use the following definition:

$$R(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T s(t) s(t + \tau) dt$$

**EXAMPLE 2.2**

A rectangular signal  $s(t)$  shows the autocorrelation function as follows (Fig. 2.7):



**FIGURE 2.7** An example of a signal and its autocorrelation function.

**2.1.3.5 The Energy (Power) Spectral Density and the Autocorrelation Function** We will show that the energy spectral density and the autocorrelation function are related as follows:

$$R(\tau) \leftrightarrow |F(\omega)|^2 \quad (2.52)$$

The definition of the autocorrelation function is rewritten as:

$$R(\tau) = \int_{-\infty}^{\infty} f(t)f(t+\tau) dt = f(-\tau) * f(\tau)$$

From the properties of the Fourier transform,  $f_1(\tau) * f_2(\tau) \leftrightarrow F_1(\omega)F_2(\omega)$  (Eq. 2.39) and  $f(-\tau) \leftrightarrow F(-\omega)$ , we have  $R(\tau) \leftrightarrow F(-\omega)F(\omega)$ . Applying  $F(-\omega) = F^*(\omega)$  (Eq. 2.37) to this, the required result is given.

For an everlasting signal, we have the following:

$$R(\tau) \leftrightarrow \lim_{T \rightarrow \infty} \frac{|F_T(\omega)|^2}{2T} \quad (2.53)$$

**2.1.3.6 Orthogonal Signals** The two signals  $s_i(t)$  and  $s_j(t)$  are orthogonal over the time range of  $t_1 - t_2$ , if

$$\int_{t_1}^{t_2} s_i(t)s_j(t)dt = 0 \quad (i \neq j)$$

Signals that never overlap in time, as shown in Figure 2.8a, are orthogonal to each other. Another group of orthogonal signals is

$$s_{cn}(t) = \cos(n\omega_0 t) \quad (n = 0, 1, 2, \dots)$$

or

$$s_{sn}(t) = \sin(n\omega_0 t) \quad (n = 0, 1, 2, \dots)$$

where  $\omega_0 = 2\pi / T_0$ ,  $T_0 = t_2 - t_1$ . Signals  $s_{cm}(t)$  and  $s_{sn}(t)$  are orthogonal even for  $m = n$ . Thus, signals  $e^{jn\omega_0 t}$  and  $e^{jm\omega_0 t}$  are also orthogonal to each other.

Orthogonal signals that take the two values  $\pm 1$  are called Walsh functions [3] and are illustrated in Figure 2.8c. The Walsh functions are given iteratively and are as follows:

$$\begin{aligned} w(0, t) &= 1, \quad 0 \leq t < 1 \\ w(1, t) &= \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & 1/2 \leq t < 1 \end{cases} \\ w(r, t) &= w\left(\left[\frac{r}{2}\right], 2t\right)w\left(r - 2\left[\frac{r}{2}\right], t\right) \end{aligned}$$

where the symbol  $\left[\frac{r}{2}\right]$  means the integer part of  $\frac{r}{2}$ .

From the properties of the Fourier transform (Eqs. 2.37 and 2.42), we have the following relation:

$$\int_{-\infty}^{\infty} f_1(t)f_2(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_1(-x)F_2(x)dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_1^*(x)F_2(x)dx$$

Thus signals whose spectra never overlap (Fig. 2.8b) are orthogonal to each other.

Another class of orthogonal signals is discussed in Appendix 7.C in the context of the orthogonal frequency division multiplexing (OFDM) systems.

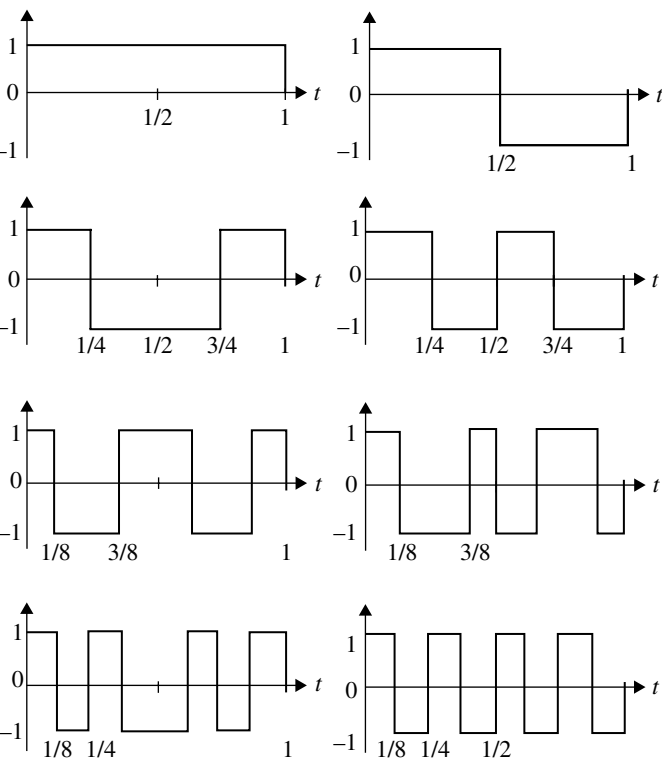
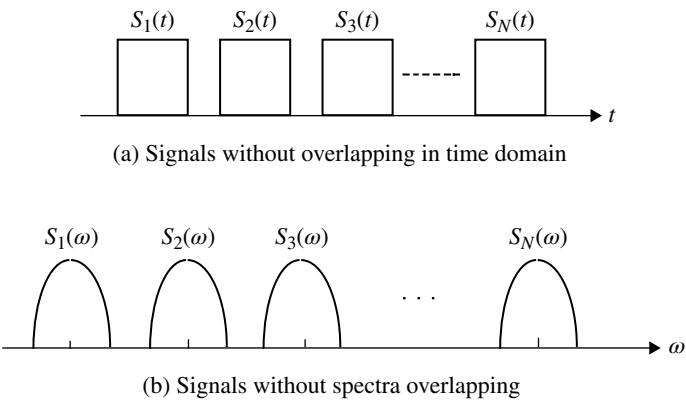
When the time region becomes infinite, the inner product,  $\int_{-\infty}^{\infty} s_i(t)s_j(t)dt$  may not be fixed in some cases, for example, the trigonometric function. However, if we modify the inner product definition as follows,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} s_i(t)s_j(t)dt$$

the result is fixed. For example, the inner product of  $e^{j\omega_i t}$  and  $e^{-j\omega_j t}$  becomes as follows:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} e^{j\omega_i t} e^{-j\omega_j t} dt = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Since  $\lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} e^{j\omega_i t} e^{-j\omega_j t} dt = 2\pi\delta(\omega_i - \omega_j)$  (Eq. 2.22), this relation is considered to show the orthogonality of signals  $e^{j\omega_i t}$  and  $e^{-j\omega_j t}$  in an infinite time region.



(c) The Walsh function

**FIGURE 2.8** Examples of orthogonal signals.

**An application of orthogonal signals:** A base station in a cellular system transmits signals for multiple users. In a digital transmission system, a signal for user  $i$  is in a form of a pulse as  $\pm s_i(t)$  ( $0 \leq t \leq T$ ), where  $s_i(t)$  is the pulse waveform and the sign “ $\pm$ ” is selected corresponding the sending data is “1” or “0.”

For  $N$  users, the transmitted signals are expressed as a whole as follows:

$$r(t) = \sum_{i=1}^N a_i s_i(t) \quad (a_i = \pm 1)$$

The signal  $r_j$  for user  $j$  is obtained by signal processing on the received signal  $r(t)$  as follows:

$$r_j = \int_0^T r(t) \cdot s_j(t) dt$$

From the orthogonality of the signals, we have:

$$r_j = \int_0^T \sum_{i=1}^N a_i s_i(t) \cdot s_j(t) dt = a_j \int_0^T s_j^2(t) dt$$

Thus, the own signal for user  $j$  is extracted and the transmitted data “1” or “0” is known by judging the sign of  $r_j$ , since  $E_b = \int_0^T s_j^2(t) dt > 0$ . Where  $E_b$  is the signal energy.

The orthogonal signals shown in Figure 2.8a, b, c are used in cellular systems, which are known as TDMA (time-division multiple access), FDMA (frequency-division multiple access), and CDMA (code-division multiple access), respectively.

## 2.1.4 Digital Signals

*Expression:* A sequence of digital signals is expressed as:

$$s(t) = \sum_{n=-\infty}^{\infty} a_n h(t - nT) \quad (2.54)$$

where  $a_n$  is a symbol that is selected from a set of discrete values corresponding to the input data signal. For example,  $a_n$  takes  $+A$  or  $-A$  for a binary system.  $h(t)$  is the impulse response of the band-limiting filter that is used to generate a pulse waveform.  $T$  denotes the symbol duration, that is, the symbol frequency is  $1/T$ . The fact that the digital signal is transmitted at a given fixed symbol frequency ensures a carrier signal regeneration and a clock timing signal regeneration at the receiver side.

If we use a non-return-to-zero (NRZ) waveform,  $h(t)$  is given as follows:

$$h(t) = \begin{cases} 1 & (0 \leq t \leq T) \\ 0 & (\text{otherwise}) \end{cases}$$

Some other waveforms are discussed in Chapter 3. Assigning the waveform to the transmitting signal is called *line coding*.

*Average Power:* The average power of a digital signal is defined as follows:

$$\begin{aligned} P &= \left\langle \lim_{N \rightarrow \infty} \frac{1}{2NT} \int_{-NT}^{NT} s^2(t) dt \right\rangle \\ &= \left\langle \lim_{N \rightarrow \infty} \frac{1}{2NT} \int_{-NT}^{NT} \left| \sum_{n=-N}^N a_n h(t-nT) \right|^2 dt \right\rangle \end{aligned}$$

where  $\langle \cdot \rangle$  denotes an ensemble average. Interchanging the operation order between the limit and the ensemble average, we have

$$P = \lim_{N \rightarrow \infty} \frac{1}{2NT} \sum_{n=-N}^N \sum_{m=-N}^N \langle a_n a_m \rangle \int_{-NT}^{NT} h(t-nT) h(t-mT) dt \quad (2.55)$$

If the occurrence of symbols is assumed to be independent and random, that is,

$$\langle a_n a_m \rangle = \begin{cases} 0 & (n \neq m) \\ \overline{a^2} & (n = m) \end{cases} \quad (2.56)$$

then Equation 2.55 yields

$$\begin{aligned} P &= \lim_{N \rightarrow \infty} \frac{\overline{a^2}}{2NT} \sum_{n=-N}^N \int_{-NT}^{NT} h^2(t-nT) dt \\ &= \frac{\overline{a^2}}{T} \int_{-\infty}^{\infty} h^2(t) dt \end{aligned} \quad (2.57)$$

Using Parseval's theorem (Eq. 2.43), the average power is rewritten as:

$$P = \frac{\overline{a^2}}{2\pi T} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega \quad (2.58)$$

Assuming an NRZ waveform, Equation 2.57 yields

$$P = \overline{a^2}$$

The same result is given by Equation 2.58, Equation 2.47 and the following formula:

$$\int_{-\infty}^{\infty} \left[ \frac{\sin x}{x} \right]^2 dx = \pi$$

The energy per symbol  $E_s$  is given as

$$E_s = PT$$

For a binary transmission, the energy per bit  $E_b$  is the same as  $E_s$ . For a  $2^n$ -ary ( $n=1,2,3,\dots$ ) transmission, it becomes

$$E_b = \frac{E_s}{n} \quad (2.59)$$

If the symbol  $a_n$  takes  $+A$  or  $-A$  (binary transmission), then

$$\overline{a_n^2} = A^2$$

For a case of four-level transmission, where  $a_n$  takes  $\pm A, \pm A/3$ , then

$$\overline{a_n^2} = \{\text{Prob}(A) + \text{Prob}(-A)\} A^2 + \left\{ \text{Prob}\left(\frac{A}{3}\right) + \text{Prob}\left(\frac{-A}{3}\right) \right\} \frac{A^2}{9}$$

If the symbols  $a_n$  are equally likely, then the probabilities are all the same, that is,  $1/4$ :

$$\overline{a_n^2} = \frac{5}{9} A^2$$

**Power Spectral Density:** The PSD of a random signal  $s(t)$  is defined as follows:

$$S(\omega) = \left\langle \lim_{N \rightarrow \infty} \frac{1}{2NT} \left| \int_{-NT}^{NT} s(t) e^{-j\omega t} dt \right|^2 \right\rangle$$

Using Equation 2.54, and changing the order of the limit and the ensemble average, in the above equation, we get:

$$S(\omega) = \lim_{N \rightarrow \infty} \frac{1}{2NT} \sum_{n=-N}^N \sum_{m=-N}^N \langle a_n a_m \rangle \int_{-NT}^{NT} \int_{-NT}^{NT} h(t-nT) e^{-j\omega t} h(x-mT) e^{j\omega x} dt dx$$

Assuming that  $a_n$  is random and using Equation 2.56, we get:

$$S(\omega) = \lim_{N \rightarrow \infty} \frac{\overline{a^2}}{2NT} \sum_{n=-N}^N \left| \int_{-NT}^{NT} h(t-nT) e^{-j\omega t} dt \right|^2 \quad (2.60)$$

Using the Fourier transform formula, the integral in the above equation becomes

$$\begin{aligned} \left| \int_{-\infty}^{\infty} h(t-nT) e^{-j\omega t} dt \right|^2 &= |H(\omega) e^{-j\omega nT}|^2 \\ &= |H(\omega)|^2 \end{aligned}$$

Using the above equation, Equation 2.60 can be reduced to

$$S(\omega) = \frac{\overline{a^2}}{T} |H(\omega)|^2 \quad (2.61)$$

Thus, the PSD of a random digital signal is found by taking the Fourier transform of a pulse waveform for the case when the data are random.

### 2.1.5 Modulated Signals

An oscillating carrier signal is expressed as:

$$\begin{aligned} c(t) &= A_0 \cos(\omega_c t + \theta_0) \\ &= \text{Re} \left( A_0 e^{j(\omega_c t + \theta_0)} \right) \end{aligned}$$

where  $A_0$  is the amplitude,  $\omega_c$  the carrier angular frequency, and  $\theta_0$  the initial phase. The carrier signal has two degrees of freedom, namely the amplitude and the phase or its derivative; the derivative of the phase is in another word the instantaneous frequency. Modulation is a process where the amplitude and/or the phase vary according to the modulating signal. Amplitude modulation is given by the following:

$$A(t) = k_A m(t)$$

where  $A(t)$  is the amplitude,  $k_A$  is a dimensionless constant, and  $m(t)$  is the modulating signal. Phase modulation is given as

$$\theta(t) = k_p m(t)$$

and frequency modulation is expressed as

$$\omega(t) = \frac{d}{dt} \theta(t) = k_f m(t)$$

where  $k_p$  and  $k_f$  are constants that have dimensions of [rad/V] and [rad/s/V], respectively. There is no difference in the above expressions for analog or digital modulation. The input  $m(t)$  can be an arbitrary modulating signal in either the analog or digital format.

If the amplitude and phase are modulated simultaneously, the modulated signal together with its carrier signal is expressed as follows:

$$s(t) = A(t) \cos[\omega_c t + \theta(t)]$$

The above equation is rewritten as:

$$s(t) = A(t) \cos \theta(t) \omega_c t - A(t) \sin \theta(t) \sin \omega_c t \quad (2.62)$$



We introduce new variables such that

$$\begin{aligned}x(t) &= A(t) \cos \theta(t) \\ y(t) &= A(t) \sin \theta(t)\end{aligned}$$

Then, Equation 2.62 is written as

$$s(t) = x(t) \cos \omega_c t - y(t) \sin \omega_c t \quad (2.63)$$

where  $x(t)$  and  $y(t)$  are called in-phase and quadrature phase components, respectively.

Because  $\cos \omega_c t$  and  $\sin \omega_c t$  are orthogonal to each other, two different modulating signals  $m_1(t)$  and  $m_2(t)$  can be loaded on the same carrier signal by letting  $x(t) = km_1(t)$  and  $y(t) = km_2(t)$ . This kind of modulation, called quadrature amplitude modulation (QAM), is a widely used digital modulation technique. This modulation is equivalent to a system where the amplitude and the phase are simultaneously modulated, such that,

$$\begin{aligned}A(t) &= \sqrt{x^2(t) + y^2(t)} \\ \theta(t) &= \tan^{-1} \frac{y(t)}{x(t)}\end{aligned}$$

The amplitude and the phase are simultaneously varied in the case of the single-side band modulation (SSB); however, this is not a QAM. For an SSB signal, there is a relation that  $x(t) = \tilde{y}(t)$ , where  $\tilde{y}(t)$  denotes the Hilbert transform; that is,  $90^\circ$  phase shifting of  $y(t)$ .

Modulated signal can be expressed as a trajectory on a two-dimensional plane of the in-phase and the quadrature phase components, which is rotating at a speed of a carrier frequency. If we observe the signal trajectory relative to the rotating plane, we need not care about the carrier frequency. In this case, using the complex plane expression, the signal can be written as follows:

$$\begin{aligned}s(t) &= A(t)e^{j\theta(t)} \\ &= x(t) + jy(t)\end{aligned}$$

This expression is sometimes called a complex amplitude, or a zero intermediate frequency (IF) signal. The actual signal is given by  $\text{Re}[s(t)e^{j\omega_c t}]$ .

For a (real) transmit signal,  $m(t)$ , the complex amplitude (base band) signals  $s(t)$  are expressed as:

$$s(t) = \begin{cases} k_A m(t) & (\text{AM}) \\ k_A \{m(t) \pm j\hat{m}(t)\} & (\text{SSB}) \\ k_A \{m_1(t) + jm_2(t)\} & (\text{QAM}) \\ A_0 \{\cos \theta(t) + j \sin \theta(t)\} & (\text{PM, FM}) \end{cases}$$

$$\text{where, } \theta(t) = \begin{cases} k_p m(t) & \text{(PM)} \\ k_F \int^t m(\tau) d\tau & \text{(FM)} \end{cases}$$

Let us investigate the spectrum  $F(\omega)$  of a signal  $f(t) = \text{Re}[s(t)e^{j\omega_c t}]$ .

Rewriting  $f(t)$  as  $f(t) = \frac{1}{2}s(t)e^{j\omega_c t} + \frac{1}{2}s^*(t)e^{-j\omega_c t}$  and using the Fourier transform properties (Eqs. 2.30 and 2.36), we have

$$F(\omega) = \frac{1}{2}S(\omega - \omega_c) + \frac{1}{2}S^*(-\omega - \omega_c) \quad (S(\omega) \leftrightarrow s(t))$$

Thus, the spectrum of a modulated signal becomes frequency-shifted by  $\omega_c$  version of the spectrum  $S(\omega)$  of the input signal.

For a linear modulation as AM, SSB, QAM, the spectrum  $S(\omega)$  of a complex base band signal is linearly expressed by the spectrum of modulating signal(s). However, for a nonlinear modulation as PM and FM, the modulated signal spectrum is expressed by a nonlinear function of the modulating signal spectrum as discussed in Section 5.4.

### 2.1.6 The Equivalent Baseband Complex Expression

When we analyze a communication system or make a computer simulation experiment with digital signal processing, the equivalent baseband complex expression of radio frequency (RF) signals is effective in view of description simplicity and a lower sampling frequency. This expression can be applied to a nonlinear as well as linear system, as long as the condition that the higher harmonic component spectra never overlap with each other is satisfied. This condition is usually satisfied in wireless communication systems. The RF signals that meet this condition are called narrowband bandpass signals.

We consider a nonlinear system, whose output is given as a series of instantaneous input value. The input signal  $x(t)$  is expressed in general as follows:

$$\begin{aligned} x(t) &= A(t) \cos[\omega_c t + \theta(t)] \quad (A(t) \geq 0) \\ &= \text{Re}[A(t)e^{j\theta} e^{j\omega_c t}] = \text{Re}[f_b(t)e^{j\omega_c t}] \end{aligned}$$

The  $n$ th order nonlinearly distorted signal is given as:

$$x^n(t) = A^n(t) \cos^n[\omega_c t + \theta(t)]$$

The fundamental component of  $A^n(t) \cos^n[\omega_c t + \theta(t)]$  becomes  $c_n A^n(t) \cos[\omega_c t + \theta(t)]$ , and it exists only for odd number  $n$ . Since  $A^n(t) = |f_b(t)|^n$ , the equivalent baseband expression becomes as follows:

$$c_n |f_b(t)|^{2m} f_b(t) \quad (n = 2m + 1)$$

where coefficients  $c_n$  are real. Some examples of  $c_n$  are  $c_3 = 3/4$ ,  $c_5 = 10/16$ , and  $c_7 = 35/64$ . If we want to express the phase shift in higher order harmonics, we use  $c_n e^{j\theta_n}$ , instead of  $c_n$ .

We consider a bandpass hard limit circuit for another example of nonlinear systems. Since the amplitude of the output signal becomes constant ( $A_0$ ), the output signal  $y(t)$  is expressed as follows:

$$y(t) = \frac{A_0 f_b(t)}{|f_b(t)|} = \frac{A_0 f_b(t)}{\sqrt{f_b^*(t) f_b(t)}}$$

When we consider a dispersive circuit that shows a non-flat-frequency response, the output signal at a given instant depends on the earlier as well as the present input signal: this circuit is called a memory system.

The power series expression of the nonlinear memory system is called the Volterra series and is given as follows [4]:

$$\begin{aligned} y(t) = & \int_{-\infty}^{\infty} h_1(\tau_1) x(t - \tau_1) d\tau_1 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2) x(t - \tau_1) x(t - \tau_2) d\tau_1 d\tau_2 \\ & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(\tau_1, \tau_2, \tau_3) x(t - \tau_1) x(t - \tau_2) x(t - \tau_3) d\tau_1 d\tau_2 d\tau_3 \\ & \dots + \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\tau_1, \tau_2, \dots, \tau_n) x(t - \tau_1) x(t - \tau_2) \dots x(t - \tau_n) d\tau_1 \dots d\tau_n \end{aligned} \quad (2.64)$$

where  $h_i(\tau_1, \tau_2, \dots, \tau_i)$  represents the nonlinear characteristics.

## 2.2 NOISE ANALYSIS

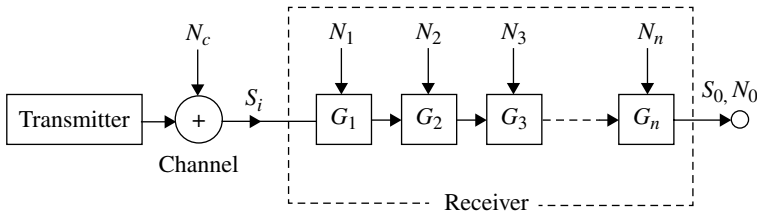
Noise is a random process that disturbs the signal transmission. In radio communications, most of the noise is added to the signal in the radio channel and at the receiver. The noise that occurs at the transmitter hardly deteriorates the signal transmission quality, because the signal level is sufficiently high.

### 2.2.1 Noise in Communication System

Figure 2.9 shows a system model for noise in a radio transmission system. The received signal level and the received noise level are denoted as  $S_i$  and  $N_c$ , respectively. Noise is generated in various stages in the circuit at the receiver. The first stage is usually a low-noise amplifier. We consider a model, where noise is assumed to be generated at the input for each stage of the circuit.

The signal to noise power ratio (SNR) at the output of the receiver is given as follows:

$$\left[ \frac{S}{N} \right]_{\text{out}} = \frac{S_i}{N_c + N_1 + \frac{N_2}{G_1} + \frac{N_3}{G_1 G_2} + \dots + \frac{N_n}{G_1 G_2 \dots G_{n-1}}}$$



**FIGURE 2.9** A model for noise in a transmission system.

where  $N_i$  and  $G_i$  ( $i = 1, 2, \dots, n$ ) are respectively the noise level and the power gain of each stage. Assuming  $G_i \gg 1$ , we can see the noise level at the stages in downstream hardly affects the output SNR. This is the reason why we place a low-noise amplifier at the first stage of the receiver. It is also seen that an effort to suppress the noise levels at a receiver much lower than the channel-induced noise level  $N_c$  is meaningless.

The noise power spectrum is usually much wider than that for the signal spectrum. Thus, in general, a flat (double-sided) noise PSD (white noise) is assumed as follows:

$$N(\omega) = \frac{N_0}{2} (-\infty < \omega < \infty) [\text{W/Hz}]$$

where we omitted suffix  $i$  for  $i$ th stage of a receiver. If we denote the power transfer function of a circuit by  $G(\omega)$ , then the noise power is given by the following:

$$N = \frac{1}{2\pi} \int_{-\infty}^{\infty} N(\omega) G(\omega) d\omega = N_0 W_{\text{eq}}$$

where,  $W_{\text{eq}}$  is an equivalent noise bandwidth, which is given as

$$W_{\text{eq}} = \frac{1}{2\pi} \int_0^{\infty} G(\omega) d\omega$$

Figure 2.10 illustrates the meaning of  $W_{\text{eq}}$ .

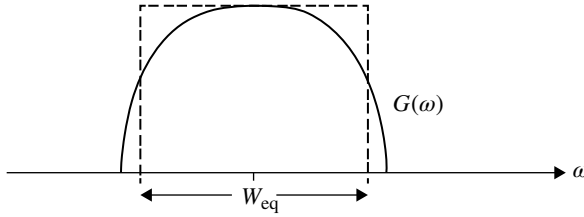
The SNR always deteriorates at the output of the circuit. The noise figure NF is defined as the degradation of SNR, for a single-sided input noise power density given by

$$N(\omega) = kT_a \quad (0 < \omega < \infty)$$

where  $k$  is the Boltzmann constant ( $k = 1.38 \times 10^{-23} \text{ (W/Hz)/K}$ ) and  $T_a$  is the temperature of the environment at which the measurement is performed. Thus, we have

$$NF = \frac{S_i / kT_a W_{\text{eq}}}{S_o / N_{\text{out}}} = 1 + \frac{T_e}{T_a} \quad (2.65)$$

where  $T_e$  is called an equivalent noise temperature and is given by



**FIGURE 2.10** Illustration of equivalent noise bandwidth.

$$T_e = \frac{N'_0}{k}$$

and  $N'_0 (= N_{\text{out}}/G, G : \text{gain})$  is an input equivalent noise power density of the circuit. In the definition of the noise figure,  $T_a = 290 \text{ K}$  is adopted.

If we place an attenuator where no noise is added before a circuit with noise temperature  $T_e$ , then the total noise figure is given by

$$NF = 1 + \frac{1}{L} \frac{T_e}{T_a}$$

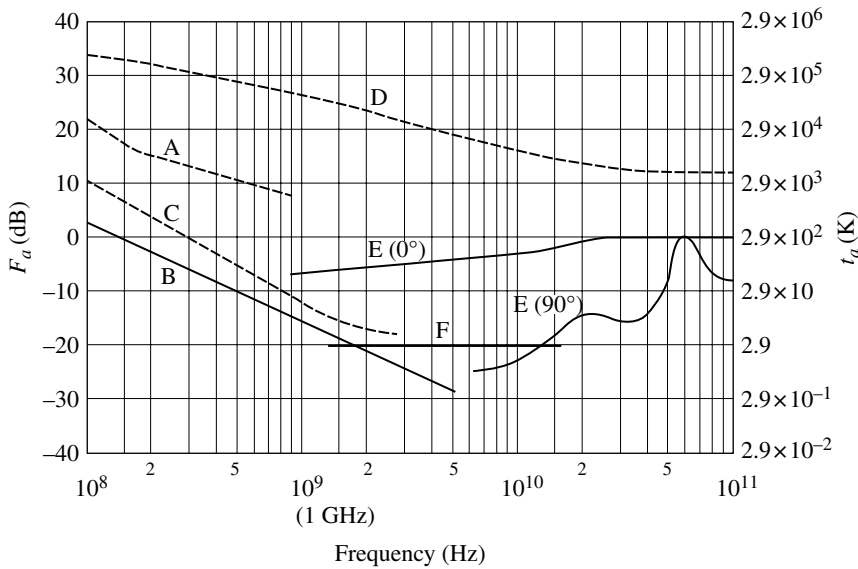
where  $L (<1)$  is the power gain of the attenuator. It is seen that an attenuator degrades the noise figure.

The external noises include man-made noise, atmospheric noise, and cosmic noise. The noise temperatures of these noise are shown in Figure 2.11 (Rec. ITU-R P.372-7). In these days, man-made noise may have increased with the evolution of electronic devices using a high frequency. The frequency region from 1 to several 10 GHz is sometimes called the radio window, since all the external noise are quite low. In this region of frequency, the effect of noise figure improvement is high and therefore the frequency in the region is used, for example, satellite communications with a low noise receiver. The cosmic microwave background radiation (F in the figure) has noise temperature of 2.7 K, which was measured (3.5 K actually measured) first in the world in a research and development of a ultra-low-noise microwave receiver. The background noise of 2.7 K had been anticipated based on a theory of the expanding universe. The engineers who measured the noise temperature were awarded the Nobel Prize in physics in 1978 for proving the theory.

### 2.2.2 Statistics of Noise

Statistical characteristics of a noise signal  $\tilde{x}$  is described by a probability distribution function  $P(x)$ . The probability distribution function  $P(x)$  is defined by the following:

$$P(x) = \text{Prob}(\tilde{x} < x) \quad (-\infty < x < \infty)$$



**FIGURE 2.11** External noise temperature (right).

where  $\text{Prob}(\tilde{x} < x)$  is a probability that  $\tilde{x}$  takes a value lower than  $x$ .  $P(x)$  is given as

$$P(x) = \int_{-\infty}^x p(x) dx$$

where  $p(x)$ , probability density function, is defined as

$$p(x) = \frac{d}{dx} P(x)$$

The probability that  $\tilde{x}$  takes a value  $x$  between  $x_1$  and  $x_2$  is given as

$$\text{Prob}(x_1 \leq \tilde{x} \leq x_2) = P(x_2) - P(x_1) = \int_{x_1}^{x_2} p(x) dx$$

In a communication system, zero-mean additive white Gaussian noise is often encountered. Its probability density function is given by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (2.66)$$

where  $\sigma^2$  is average noise power, that is,

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 p(x) dx$$

We can confirm this result, by integrating by parts and using the formula

$$\int_{-\infty}^{\infty} e^{-a^2 x^2} dx = \frac{\sqrt{\pi}}{a}$$

The mean value  $x_m$  is given by

$$x_m = \int_{-\infty}^{\infty} x p(x) dx = 0$$

In general, Gaussian noise is generated when a large number of relatively independent impulsive perturbations are applied to a low-pass filter. The effects of the impulsive perturbations are collected due to the finite impulse response time (the central limit theorem [Section 2.2.8]). This is a good model to describe how noise is generated in electronic devices, such as a transistor, where many electrons collide at random with obstacles and where the response time of the device is long enough to receive a number of collisions. From this nature of the Gaussian noise, it can be seen that the filtered Gaussian noise is still Gaussian.

The probability that a Gaussian noise takes a higher value than a given constant  $A$  is expressed as

$$\begin{aligned} \text{Prob}(x \leq A) &= \int_A^{\infty} p(x) dx \\ &= Q\left(\frac{A}{\sigma}\right) \end{aligned}$$

where

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (2.67)$$

The following function (complementary error function) is sometimes used instead of the function  $Q(x)$

$$\text{erfc}(x) = 2Q(\sqrt{2}x) \quad \left( Q(x) = \frac{1}{2} \text{erfc}\left(\frac{x}{\sqrt{2}}\right) \right) \quad (2.68)$$

where

$$\begin{aligned}\operatorname{erfc}(x) &= 1 - \operatorname{erf}(x) \\ &= \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt\end{aligned}$$

### 2.2.3 Power Spectral Density of Noise

Let  $x_T(t)$  be a noise signal given by time truncating a signal  $x(t)$  within a time range  $|t| < T$ . The PSD is defined as

$$S_x(\omega) = \lim_{T \rightarrow \infty} \frac{\langle |X_T(\omega)|^2 \rangle}{2T} \quad (2.69)$$

where  $\langle \cdot \rangle$  means ensemble average and  $X_T(\omega) \leftrightarrow x_T(t)$ , that is,

$$X_T(\omega) = \int_{-\infty}^{\infty} x_T(t) e^{-j\omega t} dt = \int_{-T}^T x(t) e^{-j\omega t} dt$$

We have

$$\begin{aligned}\langle |X_T(\omega)|^2 \rangle &= \left\langle \int_{-T}^T x(t_1) e^{j\omega t_1} dt_1 \cdot \int_{-T}^T x(t_2) e^{-j\omega t_2} dt_2 \right\rangle \\ &= \int_{-T}^T \int_{-T}^T \langle x(t_1) x(t_2) \rangle e^{j\omega(t_1 - t_2)} dt_1 dt_2\end{aligned} \quad (2.70)$$

If  $x(t)$  is a stationary process or at least  $\langle x(t_1) x(t_2) \rangle$  is not dependent on the time origin, then we can let

$$\langle x(t_1) x(t_2) \rangle = R_x(t_1 - t_2),$$

and the double integral in Equation 2.70 is given by a single integral as follows [5]:

$$\int_{-T}^T \int_{-T}^T \langle x(t_1) x(t_2) \rangle e^{j\omega(t_1 - t_2)} dt_1 dt_2 = \int_{-2T}^{2T} (2T - |\tau|) R_x(\tau) e^{-j\omega\tau} d\tau \quad (2.71)$$

where  $\tau = t_2 - t_1$ .

Substituting Equation 2.71 into Equation 2.70, we get

$$\begin{aligned}S_x(\omega) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-2T}^{2T} (2T - |\tau|) R_x(\tau) e^{-j\omega\tau} d\tau \\ &= \lim_{T \rightarrow \infty} \int_{-2T}^{2T} \left( 1 - \frac{|\tau|}{2T} \right) R_x(\tau) e^{-j\omega\tau} d\tau\end{aligned}$$



If  $\lim_{T \rightarrow \infty} \int_{-2T}^{2T} \frac{|\tau|}{2T} R_x(\tau) e^{-j\omega\tau} d\tau = 0$ , then the PSD is given as

$$S_x(\omega) = \int_{-\infty}^{\infty} R_x(\tau) e^{-j\omega\tau} d\tau$$

Thus, we have

$$S_x(\omega) \leftrightarrow R_x(\tau)$$

**2.2.3.1 Power Spectral Density of Filtered Noise** If we denote filtered noise by  $y(t)$ , then the correlation function of  $y(t)$  is given as

$$\begin{aligned} R_y(\tau) &= \langle y(t)y(t+\tau) \rangle \\ &= \langle x(t) * h(t) \rangle \langle x(t+\tau) * h(t+\tau) \rangle \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle x(t-t_1)x(t+\tau-t_2) \rangle h(t_1)h(t_2) dt_1 dt_2 \end{aligned} \quad (2.72)$$

where  $x(t)$  is input noise and  $h(t)$  is impulse response of the filter. The PSD is given by

$$S_y(\omega) = \int_{-\infty}^{\infty} R_y(\tau) e^{-j\omega\tau} d\tau \quad (2.73)$$

Inserting Equation 2.72 into Equation 2.73, and changing variables, we get

$$S_y(\omega) = |H(\omega)|^2 S_x(\omega) \quad (2.74)$$

Where,  $H(\omega) \leftrightarrow h(t)$  and  $S_x(\omega)$  are PSD of  $x(t)$  and are given as

$$S_x(\omega) = \int_{-\infty}^{\infty} R_x(\tau) e^{-j\omega\tau} d\tau \quad (2.75)$$

where  $R_x(\tau) = \langle x(t)x(t+\tau) \rangle$ .

## 2.2.4 Autocorrelation Function of Filtered Noise

From Equations 2.74 and 2.75, autocorrelation function of filtered noise is given as

$$R_y(\tau) = \langle y(t)y(t+\tau) \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 S_x(\omega) e^{j\omega\tau} d\omega$$

If the input signal is white noise with PSD of  $N_0$ , then

$$R_y(\tau) = \frac{N_0}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 e^{j\omega\tau} d\omega$$

Considering the relation given by Equation 2.52, we have

$$R_y(\tau) = N_0 \int_{-\infty}^{\infty} h(t)h(t+\tau)dt \quad (2.76)$$

It can be shown that the filtered noise shows no correlation at different symbol time ( $R_y(nT_s)=0$  for  $n \neq 0$ , where  $T_s$  is the symbol duration), if  $|H(\omega)|^2$  meets the Nyquist-I criterion (Eqs. 3.2 and 3.33). This fact means that we cannot predict a noise at a sampling instant by using noises at different sampling instants, and, therefore, cannot decrease the noise power by subtraction of the predicted noise.

### 2.2.5 Bandpass Noise

The RF noise  $n(t)$  is expressed as

$$n(t) = n_x(t) \cos \omega_c t - n_y(t) \sin \omega_c t \quad (2.77)$$

where  $n_x(t)$  and  $n_y(t)$  are stationary and independent base band noise processes that have zero mean ( $\langle n_x(t) \rangle = \langle n_y(t) \rangle = 0$ ) and  $\omega_c$  is a reference center frequency. Since  $n_x(t)$  and  $n_y(t)$  are independent, the phase of  $n(t)$  is uniformly distributed in the complex plain of  $\tilde{n}(t) = n_x(t) + jn_y(t)$ .

The amplitude of the noise is given as

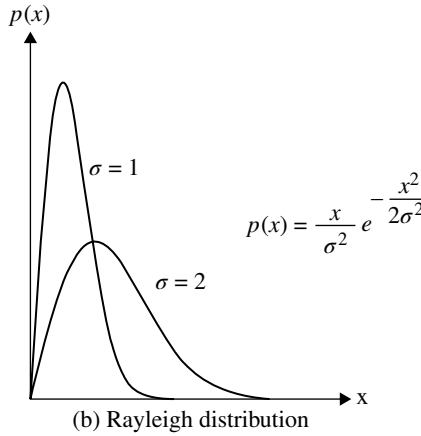
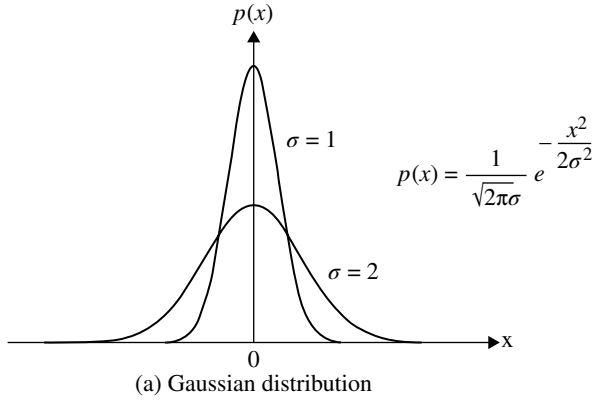
$$r(t) = \sqrt{n_x^2(t) + n_y^2(t)} \quad (2.78)$$

If we assume zero-mean Gaussian noise of the same power for  $n_x(t)$  and  $n_y(t)$ , the distribution of the amplitude obeys the Rayleigh distribution as is shown in the following. Using Equations 2.77 and 2.78, transforming the variables by  $n_x(t) = r(t)\cos\theta$  and  $n_y(t) = r(t)\sin\theta$ , and considering  $p(x,y)dxdy = p(r,\theta)rdrd\theta$ , we have

$$\begin{aligned} P(r) &= \text{Prob}(r(t) \leq r) \\ &= \int_0^{2\pi} \int_0^r \frac{1}{2\pi\sigma^2} e^{-r^2/2\sigma^2} r dr d\theta = 1 - e^{-r^2/2\sigma^2} \end{aligned}$$

The probability density function is given as

$$\begin{aligned} p(r) &= \frac{d}{dr} P(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} \quad (0 \leq r \leq \infty) \\ p(\theta) &= \frac{1}{2\pi} \quad (-\pi \leq \theta \leq \pi) \end{aligned} \quad (2.79)$$



**FIGURE 2.12** Gaussian and Rayleigh distributions.

The instantaneous power  $n^2(t)$  of  $n(t)$  changes fast with the reference frequency  $\omega_c$ . By taking the time average in duration of  $2\pi/\omega$ ,  $r^2(t)/2$  is given, where the change of  $r(t)$  in the time period is assumed to be negligibly small. The probability density  $p(q)$  of the power  $q = r^2(t)/2$  is given by letting  $p(q)dq = p(r)dr$  in Equation 2.79 as follows:

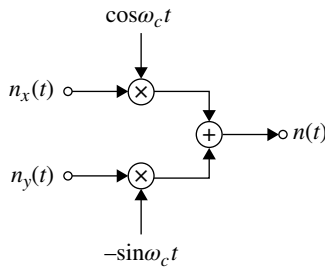
$$p(q) = \frac{1}{b} e^{-\frac{q}{b}} \quad (2.80)$$

where,  $b = \langle q \rangle = \sigma^2$ .

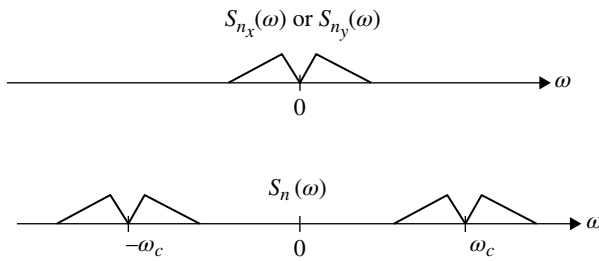
The probability density functions for the Gaussian distribution and the Rayleigh distribution are shown in Figure 2.12.

**2.2.5.1 Power Spectral Density of Passband Noise** The passband signal  $n(t)$  (Eq. 2.77) is generated as shown in Figure 2.13a. The noise PSD  $S_n(\omega)$  is given as (Eq. 2.75):

$$s_n(\omega) = \int_{-\infty}^{\infty} R_n(\tau) e^{-j\omega\tau} d\tau$$



(a) Circuit configuration



(b) Power spectrum

**FIGURE 2.13** Generation of passband noise from baseband noise.

where  $R_n(\tau)$  is the autocorrelation function of  $n(t)$ , that is,

$$R_n(\tau) = \left\langle \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T n(t+\tau)n(t)dt \right\rangle \quad (2.81)$$

Inserting Equation 2.77 into Equation 2.81 and using the relations  $\langle n_x(t)n_y(t+\tau) \rangle = \langle n_y(t) \rangle \langle n_x(t+\tau) \rangle = 0$ , we have

$$R_n(\tau) = \frac{1}{2} \{ R_x(\tau) + R_y(\tau) \} \cos \omega_c \tau \quad (2.82)$$

where  $R_x(\tau) = \langle n_x(t)n_x(t+\tau) \rangle$  and  $R_y(\tau) = \langle n_y(t)n_y(t+\tau) \rangle$  are the autocorrelation function of  $n_x(t)$  and  $n_y(t)$ , respectively.

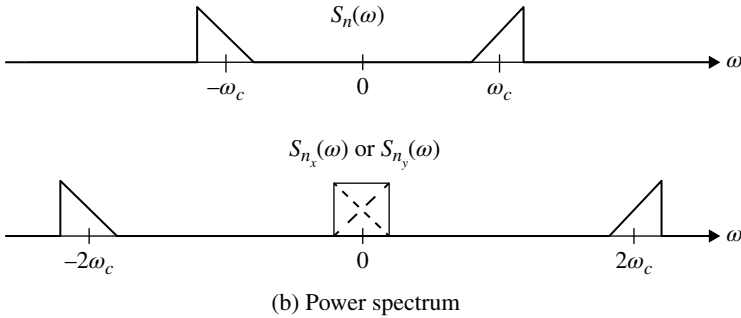
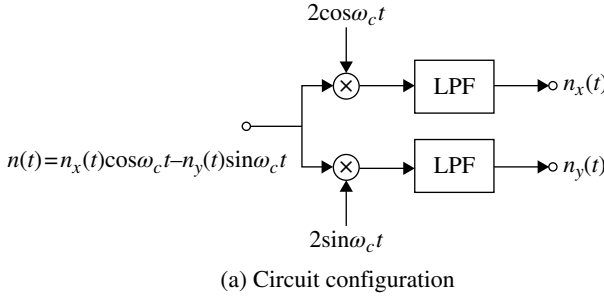
Assuming that

$$R_x(\tau) = R_y(\tau) = R_0(\tau) \quad (2.83)$$

we have

$$\langle n^2(t) \rangle = \langle n_x^2(t) \rangle = \langle n_y^2(t) \rangle$$

Rewriting  $\cos \omega_c \tau = (e^{j\omega_c \tau} + e^{-j\omega_c \tau})/2$  and from Equations 2.82 and 2.83, we have



**FIGURE 2.14** Conversion of passband noise in to quadrature baseband noise.

$$S_n(\omega) = \frac{S_0(\omega + \omega_c) + S_0(\omega - \omega_c)}{2}$$

where  $S_0(\omega) = \int_{-\infty}^{\infty} R_0(\tau) e^{-j\omega\tau} d\tau$  is the PSD of  $n_x(t)$  and  $n_y(t)$ .

Next we express the PSD of the baseband signals  $n_x(t)$  and  $n_y(t)$  with that of  $n(t)$ .  $n_x(t)$  and  $n_y(t)$  can be obtained by applying the passband signal into the circuit shown in Figure 2.14, where the role of the low-pass filter is to pass only the baseband signal. The output signal at the multipliers becomes  $m_x(t) = 2n(t)\cos\omega_c t$  and  $m_y(t) = 2n(t)\sin\omega_c t$ . Following the above argument for  $n_x(t)$  or  $n_y(t)$  instead of  $n(t)$  in Equation 2.77, we get

$$S_{n_x}(\omega) = S_{n_y}(\omega) = S_n(\omega + \omega_c) + S_n(\omega - \omega_c) \quad (2.84)$$

If we assume a narrow, flat passband spectrum, as,

$$S_n(\omega) = \begin{cases} \frac{N_0}{2} & (|\omega \pm \omega_c| \leq \omega_B, \quad \omega_B < \omega_c) \\ 0 & (\text{otherwise}) \end{cases}$$

where  $N_0/2$  is the double-sided noise power density, and then we have the baseband noise power spectrum as follows:

$$S_{n_x}(\omega) = S_{n_y}(\omega) = \begin{cases} N_0 & (|\omega| \leq \omega_B) \\ 0 & (\text{otherwise}) \end{cases}$$

In the previous discussion, the signal spectrum is considered in the frequency range,  $-\infty < \omega < \infty$ . The PSD is often defined only for positive frequency. In this case, if we consider the single-sided noise power density of  $N_0$ , the same result can be obtained in the analysis of the system.

### 2.2.6 Envelope and Phase of a Sinusoidal Signal in Bandpass Noise

Let us consider a case when a sinusoidal signal  $A \cos(\omega_c t + \phi)$  is mixed with passband Gaussian noise  $n(t)$  [5]. The mixed signal can be written as

$$\begin{aligned} z(t) &= A \cos(\omega_c t + \phi) + n_x(t) \cos(\omega_c t + \phi) - n_y(t) \sin(\omega_c t + \phi) \\ &= R(t) \cos[\omega_c t + \phi + \theta(t)] \end{aligned}$$

where the envelope  $R(t)$  and phase  $\theta(t)$  are

$$\begin{aligned} R(t) &= \sqrt{[A + n_x(t)]^2 + n_y^2(t)} \\ \theta(t) &= \tan^{-1} \frac{n_y(t)}{A + n_x(t)}, \quad |\theta| \leq \pi \end{aligned}$$

and where  $n_x(t)$  and  $n_y(t)$  are zero-mean Gaussian processes with variance  $\sigma^2$ . Dropping the notation of  $t$ , we observe

$$\begin{aligned} n_x^2 + n_y^2 &= R^2 - A^2 - 2A n_x \\ &= R^2 - 2A(A + n_x) + A^2 \\ &= R^2 - 2AR \cos \theta + A^2 \end{aligned}$$

Then, we have

$$p(R, \theta) = \frac{R}{2\pi\sigma^2} \exp \left[ \frac{-(R^2 - 2AR \cos \theta + A^2)}{2\sigma^2} \right]$$

The probability density function of the envelope is given as

$$\begin{aligned} p(R) &= \int_{-\pi}^{\pi} p(R, \theta) d\theta \\ &= \frac{R}{\sigma^2} \exp \left[ \frac{-(R^2 + A^2)}{2\sigma^2} \right] I_0 \left( \frac{AR}{\sigma^2} \right) \end{aligned} \tag{2.85}$$

where  $I_0(\cdot)$  is the modified Bessel function of the first kind.

The probability density function of the phase is given by

$$\begin{aligned} p(\theta) &= \int_0^\infty p(R, \theta) dR \\ &= \frac{1}{2\pi} e^{-A^2/2\sigma^2} \left\{ 1 + \frac{A}{\sigma} \sqrt{2\pi} \cos \theta e^{A^2 \cos 2\theta / 2\sigma^2} \left[ 1 - Q\left(\frac{A \cos \theta}{\sigma}\right) \right] \right\} \end{aligned} \quad (2.86)$$

where

$$Q(y) = \frac{1}{\sqrt{2\pi}} \int_y^\infty e^{-x^2/2} dx$$

### 2.2.7 Generation of Correlated Noises and its Probability Density Function

We generate probabilistic variables  $y_1$  and  $y_2$  with an arbitrary cross-correlation coefficient  $\rho$  and their probability density function  $h(y_1, y_2)$  through a transformation from given independent probabilistic variables  $x_1$  and  $x_2$  with probability density function  $f(x_1, x_2)$ . We assume average power never changes through the transformation of variables, that is,  $\overline{y_1^2} = \overline{x_1^2} = \sigma_1^2$  and  $\overline{y_2^2} = \overline{x_2^2} = \sigma_2^2$ . The average values are assumed to be zero for simplicity, that is,  $\overline{(x_1 = x_2)} = 0$ .

Let us assume the following transformation:

$$y_1 = x_1 \quad (2.87)$$

$$y_2 = c_1 x_1 + c_2 x_2 \quad (c_1, c_2: \text{constant}) \quad (2.88)$$

Then, from the assumptions, we have

$$\begin{aligned} \overline{y_2^2} &= c_1^2 \overline{x_1^2} + c_2^2 \overline{x_2^2} \\ &= c_1^2 \sigma_1^2 + c_2^2 \sigma_2^2 \\ &= \sigma_2^2 \end{aligned}$$

From the above equation, we get

$$c_2 = \sqrt{1 - \left( \frac{c_1 \sigma_1}{\sigma_2} \right)^2}$$

The cross-correlation coefficient  $\rho$  is defined as

$$\rho = \frac{\overline{y_1 y_2}}{\sqrt{\overline{y_1^2}} \cdot \sqrt{\overline{y_2^2}}}$$

Substituting Equations 2.87 and 2.88 in the above equation, we have  $\rho = c_1\sigma_1 / \sigma_2$ . Thus, we get

$$c_1 = \frac{\rho\sigma_2}{\sigma_1}, \quad c_2 = \sqrt{1 - \rho^2}$$

and

$$y_2 = \frac{\sigma_2}{\sigma_1} \rho x_1 + \sqrt{1 - \rho^2} x_2$$

Next, we get the probability density function  $h(y_1, y_2)$ , which is given by

$$\iint_{D_x} f(x_1, x_2) dx_1 dx_2 \rightarrow \iint_{D_y} h(y_1, y_2) dy_1 dy_2$$

where  $D_x$  and  $D_y$  show integral areas.

We express generally the relations between  $(x_1, x_2)$  and  $(y_1, y_2)$  as

$$y_1 = g_1(x_1, x_2), \quad y_2 = g_2(x_1, x_2) \quad (2.89)$$

Then, the small variations are related as

$$\begin{aligned} \Delta y_1 &= \frac{\partial g_1}{\partial x_1} \Delta x_1 + \frac{\partial g_1}{\partial x_2} \Delta x_2 = a \Delta x_1 + b \Delta x_2 \\ \Delta y_2 &= \frac{\partial g_2}{\partial x_1} \Delta x_1 + \frac{\partial g_2}{\partial x_2} \Delta x_2 = c \Delta x_1 + d \Delta x_2 \end{aligned}$$

where we put

$$\frac{\partial g_1}{\partial x_1} = a, \quad \frac{\partial g_1}{\partial x_2} = b, \quad \frac{\partial g_2}{\partial x_1} = c, \quad \frac{\partial g_2}{\partial x_2} = d$$

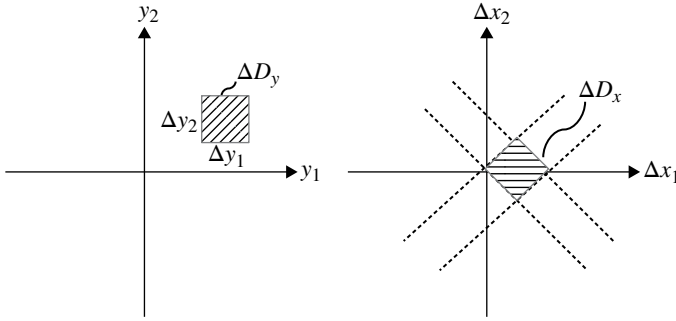
A small integral element  $\Delta D_y$  in  $(y_1, y_2)$  plain corresponds to a parallelogram  $\Delta D_x$  in  $(x_1, x_2)$  plain as shown in Figure 2.15.

If we denote the area size of  $\Delta D_x$  by  $S$ , we have  $S = \Delta y_1 \Delta y_2 / |J|$ . Where  $J$  is called the Jacobian expressed as

$$J = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} = \frac{\partial g_1}{\partial x_1} \frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2} \frac{\partial g_2}{\partial x_1} = ad - bc$$

The transformation of integral element becomes as  $dx_1 dx_2 \rightarrow \frac{1}{|J|} dy_1 dy_2$ .





**FIGURE 2.15** Transform of integral elements.

By solving Equation 2.89 with respect to  $x_1$  and  $x_2$ , we express

$$x_1 = g'_1(y_1, y_2), \quad x_2 = g'_2(y_1, y_2)$$

Then, the probability density function of  $y_1$  and  $y_2$  becomes

$$h(y_1, y_2) = f(g'_1(y_1, y_2), g'_2(y_1, y_2)) \frac{1}{|J|}.$$

*Example for Gaussian Noise* We consider a case when the powers of variables and  $x_1, x_2$  are the same, that is,  $(\sigma_1 = \sigma_2 = \sigma)$ . Using the following expressions,

$$f(x_1, x_2) = f(x_1)f(x_2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x_1^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x_2^2}{2\sigma^2}},$$

$$x_1 = y_1, \quad x_2 = \frac{y_2 - \rho y_1}{\sqrt{1 - \rho^2}}, \quad \text{and} \quad |J| = \sqrt{1 - \rho^2},$$

we have

$$h(y_1, y_2) = \frac{1}{2\pi\sigma^2\sqrt{1 - \rho^2}} e^{-\frac{y_1^2 - 2\rho y_1 y_2 + y_2^2}{2\sigma^2(1 - \rho^2)}}$$

### 2.2.8 Sums of Random Variables and the Central Limit Theorem

First, we consider two independent random variables,  $x_1$  and  $x_2$  with probability density functions  $p_1(x_1)$  and  $p_2(x_2)$ , respectively. Make a sum of them as  $x = x_1 + x_2$ . Then, the probability function  $p(x)$  is given by convolution as

$$\begin{aligned} p(x) &= \int_{-\infty}^{\infty} p_1(x_1) p_2(x - x_1) dx_1 \\ &= p_1(x) * p_2(x) \end{aligned}$$

Similarly, for a sum of  $n$  independent random variables,  $x = \sum_{i=1}^n x_i$ , we have

$$p(x) = p_1(x) * p_2(x) * \dots * p_n(x)$$

To perform the convolution,  $n-1$  times integrations are necessary. If the Fourier transform of  $p_i(x)$  is given as  $\psi_i(v) = \int_{-\infty}^{\infty} e^{jvx} p_i(x) dx$ , from the property of the Fourier transform, the probability density function is obtained by a single integration (the inverse Fourier transform) as

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \prod_{i=1}^n \psi_i(v) e^{-jvx} dv$$

Here, in convention, the sign of the exponent is opposite to usual expression of the Fourier transform. The term  $\psi_i(v)$  called the characteristic function can be seen as a statistical average of  $e^{jvx}$ , that is,  $\psi_i(v) = \langle e^{jvx} \rangle$ .

If all of the density functions are the same, we get  $\prod_{i=1}^n \psi_i(v) = [\psi_0(v)]^n$ .

In the following, we assume that the identically distributed independent random variables  $x_i$  take a zero mean ( $\langle x_i \rangle = 0$ ) and a variance  $\langle x_i^2 \rangle = \sigma^2$ . We consider a sum of the random variables as

$$x = \sum_{i=1}^n \frac{x_i}{\sqrt{n}}$$

The mean value of  $x$  is zero and the variance is  $\sigma^2$  still. The probability distribution function of  $p(x_i/\sqrt{n})$  should be given such that  $\int_{-\infty}^{\infty} p(x_i/\sqrt{n}) dx_i = 1$ . To meet this condition, the characteristic function  $\psi_i(v/\sqrt{n}) \leftrightarrow p(x_i/\sqrt{n})$  becomes

$$\psi_i\left(\frac{v}{\sqrt{n}}\right) = \int_{-\infty}^{\infty} e^{jvx/\sqrt{n}} p_i(x) dx$$

where we can confirm that  $\int_{-\infty}^{\infty} p(x_i/\sqrt{n}) dx_i = \psi_i(0) = 1$ . Then, the characteristic function of  $x$  is given as

$$\Psi(v) = \left[ \psi_0\left(\frac{v}{\sqrt{n}}\right) \right]^n$$

where  $\psi_0(v/\sqrt{n})$  is the characteristic function of  $x_i/\sqrt{n}$  ( $i=1,2,\dots,n$ ).

We take the Taylor series expansion of  $\psi_0(v/\sqrt{n})$  as

$$\psi_0\left(\frac{v}{\sqrt{n}}\right) = 1 + \frac{v}{\sqrt{n}} \frac{d}{d\omega} \psi_0(\omega) \Big|_{\omega=0} + \frac{1}{2} \frac{v^2}{n} \frac{d^2}{d\omega^2} \psi_0(\omega) \Big|_{\omega=0} + \dots$$

where  $\omega = (v/\sqrt{n})$  and  $\psi_0(0) = 1$  are used.

From the property of the Fourier transform (Eq. 2.35), we have

$$\left. \begin{aligned} \frac{d}{d\omega} \psi_0(\omega) \\ \frac{d^2}{d\omega^2} \psi_0(\omega) \end{aligned} \right|_{\omega=0} &= \int_{-\infty}^{\infty} jx p(x) dx = 0 \quad (\langle x \rangle = 0) \\ &= \int_{-\infty}^{\infty} -x^2 p(x) dx = -\sigma^2$$

Thus, we get

$$\psi_0\left(\frac{v}{\sqrt{n}}\right) \approx 1 - \frac{1}{2} \frac{v^2}{n} \sigma^2 \quad \left(\frac{v}{\sqrt{n}} \ll 1\right)$$

Taking logarithm of  $\psi(v)$ , we get

$$\begin{aligned} \ln \psi(v) &= n \ln \left[ \psi_0\left(\frac{v}{\sqrt{n}}\right) \right] \\ &\approx n \ln \left( 1 - \frac{1}{2} \frac{v^2}{n} \sigma^2 \right) \\ &\approx -\frac{1}{2} (\sigma v)^2 \quad (\ln(1+x) \approx x, \text{ for } x \ll 1) \end{aligned}$$

Therefore,  $\psi(v) = e^{-(\sigma v)^2/2}$ . Taking the inverse Fourier transform, the probability density function becomes Gaussian distribution,  $p(x) = \frac{1}{\sqrt{2b\sigma}} e^{-\frac{x^2}{2\sigma^2}}$  (Eq. 2.48).

The result that the sum of many independent random variables tends to obey the Gaussian distribution is called the central limit theorem. Note that no specific probability density function is assumed except for the assumption of a zero mean, independent and identical distribution. If we consider a variable  $x = \sum_{i=1}^n \frac{x_i}{n}$ , we get  $\ln \Psi(v) = 0$  and therefore  $p(x) = \delta(x)$  or equivalently  $\sigma \rightarrow 0$  in the Gaussian distribution.

### EXAMPLE 2.3

Probability density function for a uniform distribution with zero mean and variance of  $\sigma^2/n$  is expressed as

$$p_n(x) = \frac{\sqrt{n}}{2\sqrt{3}\sigma} \left( \frac{-\sqrt{3}\sigma}{\sqrt{n}} \leq x \leq \frac{\sqrt{3}\sigma}{\sqrt{n}} \right), \quad p_n(x) = 0 \quad (\text{otherwise})$$

Then, the characteristic function  $P_n(v)$  becomes

$$P_n(v) = \int_{-x_m}^{x_m} \frac{\sqrt{n}}{2\sqrt{3}\sigma} e^{jvx} dx = \frac{\sin(\sqrt{3/n}\sigma v)}{\sqrt{3/n}\sigma v}$$

where  $x_m = \sqrt{3/n}\sigma$ .

For a sum of random variables,  $x = \sum_{i=1}^n \frac{x_i}{\sqrt{n}}$  where  $x_i$  obeys the above uniform distribution, the characteristic function becomes

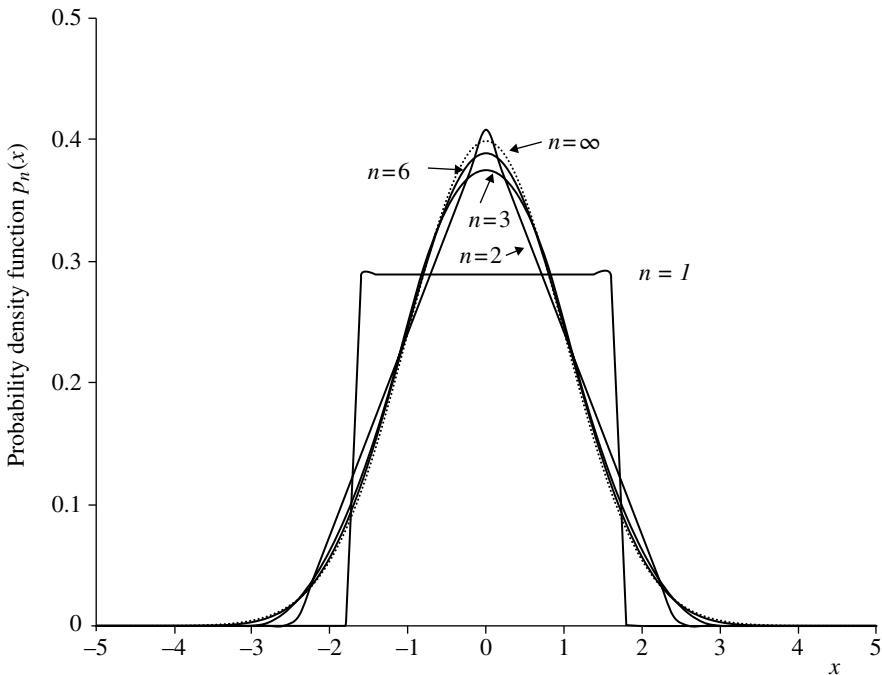
$$P_x(v) = \left[ \frac{\sin(\sqrt{3/n}\sigma v)}{\sqrt{3/n}\sigma v} \right]^n$$

The probability density function is given as

$$\begin{aligned} p(x) &= \int_{-\infty}^{\infty} P_x(v) e^{-jvx} dv \\ &= \int_{-\infty}^{\infty} \left[ \frac{\sin(\sqrt{3/n}\sigma v)}{\sqrt{3/n}\sigma v} \right]^n \cos(vx) dv \end{aligned}$$

The probability density functions for  $n=1, 2, 3$ , and  $10$  are shown in Figure 2.16. The Gibbs phenomena are observed for  $n=1$  and  $2$ .

We show the characteristic function  $P_x(v)$  approaches to a Gaussian function for  $n \rightarrow \infty$  in the following.



**FIGURE 2.16** Probability density functions given by the inverse Fourier transform.

Using the approximation,  $\sin(x) \approx x - \frac{1}{3!}x^3$  ( $|x| \ll 1$ ), we have

$$\frac{\sin(\sqrt{3/n}\sigma v)}{\sqrt{3/n}\sigma v} \approx 1 - \frac{1}{2} \frac{(\sigma v)^2}{n} \quad \text{for } n \gg 1.$$

Using  $\lim_{n \rightarrow \infty} (1 - x/n)^n = e^{-x}$ , we have

$$P_x(v) \approx \left[ 1 - \frac{1}{2} \frac{(\sigma v)^2}{n} \right]^n = e^{-\frac{\sigma^2 v^2}{2}} \quad \text{for } n \rightarrow \infty$$

## 2.3 LINEAR SYSTEM

Signal processing circuits are described with an emphasis on continuous time linear systems.

### 2.3.1 Linear Time-Invariant System

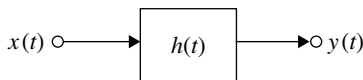
A linear system outputs a signal produced through a linear operation on an input signal (Fig. 2.17). In other words, it is a system where the principle of linear superposition holds. More specifically it is defined as follows. Consider a system that outputs signals  $y_i(t)$  ( $i=1,2$ ) for input signals  $x_i(t)$ , respectively. Then for an input signal given with a linearly weighted superposition of  $x_i(t)$  as  $x(t) = a_1 x_1(t) + a_2 x_2(t)$  ( $a_1, a_2$ : constant), its output signal becomes  $y(t) = a_1 y_1(t) + a_2 y_2(t)$ .

The time-invariant system is defined as a system where the relation between an input and an output signals never depends on the time origin. Specifically, for a delayed input signal  $x(t-t_0)$  ( $t_0$ : time delay), its output signal is also delayed as  $y(t-t_0)$ . In this book, time-invariant systems are assumed unless otherwise stated.

### 2.3.2 Response of Linear System

We obtain generally the input–output relation of a linear system. An input signal  $x(t)$  is approximated with step approximation as

$$x(t) \cong \sum_n x(n\Delta t) p(t - n\Delta t)$$



**FIGURE 2.17** Linear system.

where

$$p(t) = \begin{cases} 1 & \left( |t| \leq \frac{\Delta t}{2} \right) \\ 0 & \text{(otherwise)} \end{cases}$$

The approximation becomes precise for a smaller  $\Delta t$ . Let us denote the output signal by  $q(t)$  for an input signal  $p(t)$ . From the linearity and time-invariant property of the system, the output signal  $y(t)$  becomes  $\sum_n x(n\Delta t)q(t - n\Delta t)$  for an input signal  $\sum_n x(n\Delta t)p(t - n\Delta t)$ . If we put  $p(t)' = p(t)/\Delta t$ ,  $p(t)'$  becomes the delta function as  $\Delta t \rightarrow 0$ . The response for  $p(t)'$  is  $q(t)/\Delta t$ . Thus we have

$$\begin{aligned} y(t) &= \sum_n x(n\Delta t)q(t - n\Delta t)\Delta t / \Delta t \\ &= \int_{-\infty}^{\infty} x(\tau)h(t - \tau)d\tau \quad (\Delta t \rightarrow 0) \\ &= x(t) * h(t) \end{aligned} \quad (2.90)$$

where  $h(t) \left( = \lim_{\Delta t \rightarrow 0} q(t) / \Delta t \right)$  is called as the impulse response. The symbol “\*” means the convolution integral. If we replace the variable  $t - \tau$  with  $\tau$ , we get another expression as

$$y(t) = \int_{-\infty}^{\infty} x(t - \tau)h(\tau)d\tau \quad (2.91)$$

We can confirm that  $y(t) = h(t)$  for  $x(t) = \delta(t)$ .

If we use integral expression of a signal as  $x(t) = \int_{-\infty}^{\infty} x(\tau)\delta(t - \tau)d\tau$ , the abovementioned argument becomes simple as follows: From the linearity and time independence of the system, for an input signal  $x(\tau)\delta(t - \tau)$  we have an output  $x(\tau)h(t - \tau)$ . Therefore, for the input signal  $x(t)$  we get the output  $y(t) = \int_{-\infty}^{\infty} x(\tau)h(t - \tau)d\tau$ .

For a system to be physically realizable, its impulse response must be a real function. Although complex impulse responses are sometimes used, it is a mathematical expansion to treat two impulse responses at the same time.

**2.3.2.1 Frequency Response** The Fourier transform of an impulse response  $h(t)$  is given as

$$H(\omega) = \int_{-\infty}^{\infty} h(t)e^{-j\omega t} dt,$$

and is called as the (voltage) transfer function or the system function. We call the function  $|H(\omega)|^2$  as the power transfer function.

Let us consider an input signal  $x(t) = A_0 e^{j\omega_0 t}$  ( $-\infty \leq t \leq \infty$ ) (*alternating current-circuit theory*). From Equation 2.91, we have

$$\begin{aligned}
 y(t) &= A_0 e^{j\omega_0 t} \int_{-\infty}^{\infty} h(\tau) e^{-j\omega_0 \tau} d\tau \\
 &= A_0 e^{j\omega_0 t} H(\omega_0) \\
 &= H(\omega_0) x(t).
 \end{aligned}$$

Thus, for the special input signal, we obtain the output by multiplying the transfer function  $H(\omega_0)$  to the input signal. In the language of integral equations, we call the input signal  $x(t) = A_0 e^{j\omega_0 t}$  ( $-\infty \leq t \leq \infty$ ) as an *eigen function* and  $H(\omega_0)$  as the *eigen value*.

Since an impulse response  $h(t)$  and the transfer function are in relation of a Fourier transform pair, the system characteristics are known by either one of them. Since the delta function is generated by adding trigonometric functions with all of the frequencies (Section 2.1.1), the measurement of an impulse response  $h(t)$  is equivalent to measuring the transfer function  $H(\omega)$  for all the frequencies simultaneously. We need a narrow input pulse to measure an impulse response. In geological surveys, for example, to exploit oil resources, the narrow pulse is generated through a dynamite explosion. In RF circuit characterizations, the transfer function  $H(\omega)$  is usually measured since a measurement of the impulse response with a narrow pulse becomes difficult. Using sinusoidal signals as the input and changing its frequency continuously, amplitudes and phases (or group delays) of the output signal are measured as a function of the frequency. Equipment for that purpose is known as a network analyzer, which is indispensable in an experimental characterization of RF circuits.

A complex number expression of input signal  $x(t)$  and output signal  $y(t)$  is used because of simplicity in mathematical treatments. The actual signal is expressed with the real or imaginary part of the complex signal. Here, we consider  $x(t) = \cos \omega_0 t$  as an example. From the fact that  $H(-\omega) = H^*(\omega)$  (Eq. 2.37) for a real function, and linearity of the circuit, we have

$$\begin{aligned}
 y(t) &= \frac{1}{2} \left[ H(\omega_0) e^{j\omega_0 t} + H(-\omega_0) e^{-j\omega_0 t} \right] \\
 &= \text{Re} \left[ H(\omega_0) e^{j\omega_0 t} \right]
 \end{aligned}$$

And similarly, for  $x(t) = \sin \omega_0 t$ , we have  $y(t) = \text{Im} \left[ H(\omega_0) e^{j\omega_0 t} \right]$ .

By taking the Fourier transform of Equation 2.90 or 2.91, and from the property of the Fourier transform (Eq. 2.39), we have

$$Y(\omega) = H(\omega) X(\omega) \quad (2.92)$$

where  $X(\omega) \leftrightarrow x(t)$  and  $Y(\omega) \leftrightarrow y(t)$ .

$|H(\omega)|$  and  $\angle H(\omega) \equiv \theta(\omega)$  are called the amplitude characteristic and phase characteristic, respectively. The group delay is defined as  $\tau(\omega) = -d/d\omega \theta(\omega)$ . The physical meaning of the group delay becomes easy to understand by analogy to the relation  $\omega(t) = d\theta(t)/dt$  between instantaneous (angular) frequency  $\omega(t)$  and phase  $\theta(t)$ . If the instantaneous frequency takes a fixed value ( $\omega_0$ ), we have  $\theta(t) = \omega_0 t$ . If the time delay is the same,  $\tau(\omega) = t_0$  for all frequencies, we get  $\angle H(\omega) = -\omega t_0$ .

From Equation 2.92, we understand that  $Y(\omega_0) = 0$  for  $X(\omega_0) = 0$ , in other words, frequency components that never exist in the input signal will never appear newly in the output signal. This fact is true only for linear time-invariant systems and is not valid for systems without the linearity or time-invariant property. Some examples are shown as follows:

(i) Nonlinear time-invariant systems

$$y(t) = ax(t) + c \quad (a, c : \text{constant}), \quad y(t) = ax^2(t), \quad y(t) = a \cos x(t)$$

(ii) Linear time-variant system

$$y(t) = a(t) x(t)$$

(iii) Nonlinear time-variant system

$$y(t) = a(t) x(t) + c(t)$$

Nonlinear systems or nonlinear signal processing plays an important role in engineering field. In contrast to linear systems, we could not develop general arguments on nonlinear systems, since the behaviors strongly depend on each structural nonlinearity. An example characteristic of the nonlinear system is that the chaos phenomena appear only in a nonlinear system.

**2.3.2.2 Causality** Causality means that output (result) never happens before the input (cause) is applied. In other words, if input signal  $x(t)$  is zero for  $t < t_1$ , then the output signal  $y(t)$  is also zero for  $t < t_1$ . Describing this with Equation 2.90, we have

$$y(t) = \int_{t_1}^{\infty} h(t - \tau) x(\tau) d\tau = 0 \quad (t < t_1)$$

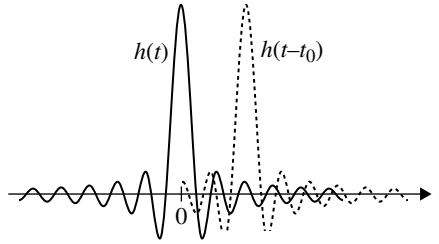
From the above equation, we get

$$h(t) = 0 \quad (t < 0)$$

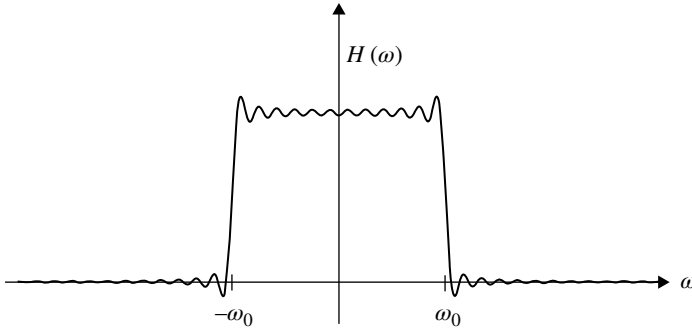
This result expresses merely the causality for a special case where the input signal is an impulse (the delta function). The first-order low-pass filter shown in Figure 2.22 has an impulse response with Equation 2.99 and fulfills the causality.

Any system without the causality is realizable. Sometimes, we theoretically treat a system that never satisfies the causality. In actual system, we shift the impulse response so that  $h(t - t_0) = 0$  ( $t < 0$ ) (Fig. 2.18). Some systems show an impulse response spanning in an infinite time region. In this case, we should approximate the impulse response by truncating it in a finite time. If we take a higher time shift  $t_0$ , the response time delay becomes longer, although the approximation becomes better. Figure 2.19 shows frequency transfer characteristics of an ideal low-pass filter with truncations of the impulse response. The ripple in the frequency transfer function never disappears even when the truncation time approaches infinite (although the





**FIGURE 2.18** Impulse response and the causality.



**FIGURE 2.19** Frequency transfer function of an ideal low-pass filter with truncation of its impulse response within  $\pm 6T$ ,  $T = 2\pi / \omega_0$ .

ripple duration tends to zero). The similar result in time domain is known as Gibbs' phenomenon.

The response of a system that satisfies the causality is given as

$$y(t) = \int_{-\infty}^t h(t-\tau)x(\tau)d\tau$$

The output  $y(t)$  is affected by  $x(t)$  in the time before  $t$ . Furthermore, if  $x(t) = 0$  ( $t \leq 0$ ) we have

$$y(t) = \int_0^t h(t-\tau)x(\tau)d\tau$$

**2.3.2.3 Stability** A system is defined as stable if the output signal is bounded, that is,  $|y(t)| < \infty$  for any bounded input signal  $x(t)$  ( $|x(t)| < \infty$ ).

A linear system is stable if and only if (necessary and sufficient) the impulse response satisfies the following condition:

$$\int_{-\infty}^{\infty} |h(t)| dt < \infty$$

*Proof: Sufficient condition*

Taking the absolute value of Equation 2.91, we can show

$$\begin{aligned} |y(t)| &= \left| \int_{-\infty}^{\infty} x(t-\tau)h(\tau)d\tau \right| \\ &\leq \int_{-\infty}^{\infty} |x(t-\tau)||h(\tau)|d\tau \leq M \int_{-\infty}^{\infty} |h(\tau)|d\tau < \infty \end{aligned}$$

where  $M$  is the maximum value of  $|x(t)|$ .

*Necessary condition*

Let us assume the condition is not necessary (proof by contradiction), that is, a system is stable even when  $\int_{-\infty}^{\infty} |h(t)|dt = \infty$ . We consider a bounded input signal as  $x(t) = h^*(-t)/|h(-t)|$ , and we have

$$y(t) = \int_{-\infty}^{\infty} \frac{h^*(-t+\tau)}{|h(-t+\tau)|} h(\tau)d\tau.$$

For  $t=0$ , we can show the system is not stable as

$$y(0) = \int_{-\infty}^{\infty} \frac{h^*(\tau)}{|h(\tau)|} h(\tau)d\tau = \int_{-\infty}^{\infty} |h(\tau)|d\tau = \infty$$

Thus, the assumption is denied and therefore the condition is necessary.

A system with its impulse  $h(t) = \sum_{i=1}^N h_i(t)$  is stable if  $\int_{-\infty}^{\infty} |h_i(t)|dt < \infty$  ( $i = 1, 2, \dots, N$ ).

*Proof:* we can show that

$$\int_{-\infty}^{\infty} |h(t)|dt \leq \sum_{i=1}^N \int_{-\infty}^{\infty} |h_i(t)|dt < \infty$$

## EXAMPLE 2.4

We consider a system with its impulse response given as

$$h(t) = \begin{cases} \frac{t^n}{n!} e^{-at} & t > 0 \\ 0 & t < 0 \end{cases} \quad (a: \text{real constant}, n = 0, 1, 2, \dots) \quad (2.93)$$

For  $n=0$ , since  $\lim_{T \rightarrow \infty} \int_0^T |h(t)|dt = \lim_{T \rightarrow \infty} \frac{1-e^{-aT}}{a}$ , the system is stable if  $a > 0$ .

The system shown in Figure 2.22 corresponds to this case and it is stable.

For  $n=1$ , performing integral by part, we have

$$\lim_{T \rightarrow \infty} \int_{-\infty}^{\infty} |h_i(t)|dt = \lim_{T \rightarrow \infty} \left\{ \left[ t \frac{e^{-at}}{-a} \right]_0^T - \int_0^T \frac{e^{-at}}{-a} dt \right\}$$

Thus, this system is stable if  $a > 0$ . Similarly, for  $n \geq 2$ , the system is stable if  $a > 0$ . If  $a$  is a complex value, the system is stable if the real part of  $a$  is positive.

Applying frequency differentiation property (Eq. 2.35) of the Fourier transform to Equation 2.93 for  $n=0$ , we have the transfer function of the system as

$$H(\omega) = \frac{1}{(a + j\omega)^{n+1}} \quad (2.94)$$

**EXAMPLE 2.5**  $h(t) = \sum_{i=1}^N b_i h_i(t)$

Where  $b_i$  and  $a_i$  are real and  $h_i(t) = \begin{cases} e^{-a_i t} & t > 0 \\ 0 & \text{otherwise} \end{cases}$ .

This system is stable if  $a_i > 0 (i = 1, 2, \dots, N)$ . Taking the Fourier transform, we have

$$H(\omega) = \sum_{i=1}^N \frac{b_i}{j\omega + a_i}.$$

Letting  $s = j\omega$  we get

$$H(s) = \sum_{i=1}^N \frac{b_i}{s + a_i}.$$

The above equation is rewritten as

$$\begin{aligned} H(s) &= \frac{\sum_{i=1}^N b_i \prod_{n=1, n \neq i}^N (s + a_n)}{\prod_i^N (s + a_i)} \\ &= \frac{\sum_{i=1}^{N-1} \beta_i s^{N-1-i}}{\sum_{i=0}^N \alpha_i s^{N-i}} = \frac{B(s)}{A(s)} \end{aligned}$$

where  $\alpha_i$  and  $\beta_i$  are constants given by expanding the products.  $A(s)$  and  $B(s)$  are polynomials of  $N$ th- and  $(N-1)$ th order, respectively.

We will investigate the stability when a transfer function  $H(s) = B(s)/A(s)$  is given beforehand. We denote the poles of  $H(s)$ , that is, the roots  $s$  for  $A(s) = 0$  by  $z_{pi}$ . From the arguments so far, the system is stable if  $\text{Re}(z_{pi}) < 0$ . This is true even if  $A(s) = 0$  has  $n$ -tuple roots as seen from the result of Example 2.4.

If  $z_{pi}$  is a root, the conjugate of the root  $z_{pi}^*$  is also a root, since the coefficients of  $H(s)$  are real. If a root is expressed as  $z_{pi} = p + jq (p, q : \text{real})$ , the impulse response  $h_i(t)$  which corresponds to the root becomes

$$h_i(t) = \begin{cases} e^{pt} e^{jq t} & t > 0 \\ 0 & t < 0 \end{cases}$$

Since the impulse response is a complex function, it is not realizable. By adding an impulse response  $h_i^*(t)$  corresponding to  $z_{pi}^* = p - jq$ , we have

$$g_i(t) = h_i(t) + h_i^*(t) = \begin{cases} 2e^{pt} \cos qt & t > 0 \\ 0 & t < 0 \end{cases}$$

Thus, the combined impulse response  $g_i(t)$  is realizable. This system is stable if  $p < 0$ . Letting  $g_i(t) \leftrightarrow G_i(\omega)$ , we get

$$G_i(\omega) = \frac{1}{j(\omega - q) - p} + \frac{1}{j(\omega + q) - p}$$

A system with the above transfer function is called as first-order bandpass filter with a center frequency of  $q$ .

### EXAMPLE 2.6

We find impulse response and investigate the stability of a system with transfer function given as

$$\begin{aligned} H(\omega) &= \frac{1}{|a|^2 + j\omega(a + a^*) + (j\omega)^2} \\ &= \frac{1}{(a + j\omega)(a^* + j\omega)} \end{aligned}$$

Since the coefficients of the polynomial of  $j\omega$  are real, the system is realizable.

$a \neq a^*$ : we modify the transfer function as

$$H(\omega) = \frac{1}{a^* - a} \frac{1}{a + j\omega} + \frac{1}{a - a^*} \frac{1}{a^* + j\omega}$$

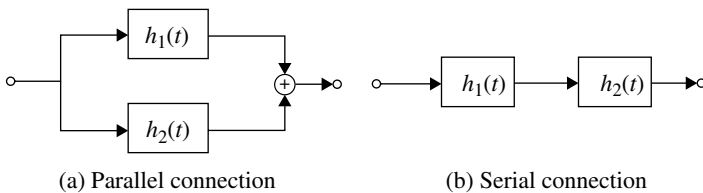
Then, we have

$$h(t) = \begin{cases} \frac{1}{a^* - a} (e^{-at} - e^{-a^*t}) & t > 0 \\ 0 & t < 0 \end{cases}$$

Putting  $a = p + jq$  ( $p, q$ : real), we obtain

$$h(t) = \begin{cases} \frac{1}{q} e^{-pt} \sin qt & t > 0 \\ 0 & t < 0 \end{cases}$$

Since  $\int_{-\infty}^{\infty} |h(t)| dt = \int_{-\infty}^{\infty} \left| \frac{1}{q} e^{-pt} \sin qt \right| dt < \frac{1}{|q|} \int_0^{\infty} e^{-pt} dt$ , the system is stable if  $p = \text{Re}\{a\} > 0$ .



**FIGURE 2.20** Connection of two systems.

(ii)  $a = a^*$ : We have

$$H(\omega) = \frac{1}{(a + j\omega)^2}$$

From Equations 2.93 and 2.94, we get  $h(t) = \begin{cases} te^{-at} & t > 0 \\ 0 & t < 0 \end{cases}$ . Then, we have

$$\begin{aligned} \int_{-\infty}^{\infty} |h(t)| dt &= \int_{-\infty}^{\infty} te^{-at} dt \\ &= \left[ t \frac{e^{-at}}{-a} \right]_0^{\infty} - \int_{-\infty}^{\infty} \frac{e^{-at}}{-a} dt \\ &= \left[ -t \frac{e^{-at}}{a} \right]_0^{\infty} - \left[ \frac{e^{-at}}{-a} \right]_0^{\infty} = \begin{cases} < \infty & a > 0 \\ \infty & a < 0 \end{cases} \end{aligned}$$

The system is stable if  $a > 0$ .

**2.3.2.4 Interconnection of Linear Systems** We consider parallel and serial (tandem) connections of two linear systems (Fig. 2.20). In the parallel connection, the impulse response and the transfer function as a whole system become the sum of them for each system, that is,  $h(t) = h_1(t) + h_2(t)$  and  $H(\omega) = H_1(\omega) + H_2(\omega)$ . In the serial connection, we can show that  $h(t) = h_1(t) * h_2(t)$  and  $H(\omega) = H_1(\omega)H_2(\omega)$ . A change of order of the connection never affects the characteristics of the connected system. However, this is not true if one of them is a nonlinear system.

### 2.3.3 System Description with Differential Equations

Systems are characterized with a differential equation regardless of their strict or approximated descriptions. The solution for a differential equation always exists and is uniquely obtained for a given initial condition (the existence and uniqueness theorem). Owing to this fact, any method is applicable to find the solution and various methods are known depending on the types of differential equations. Generally it is difficult to solve analytically nonlinear differential equations, and numerical methods are usually used. However, linear differential equations especially with constant



From the above equation, we get

$$\begin{aligned} Y(\omega) &= \frac{\sum_{n=0}^N a_n(j\omega)^n}{\sum_{m=0}^M b_m(j\omega)^m} X(\omega) \\ &= H(\omega)X(\omega) \end{aligned}$$

where

$$H(\omega) = \frac{\sum_{n=0}^N a_n(j\omega)^n}{\sum_{m=0}^M b_m(j\omega)^m} \quad (2.97)$$

is the transfer function.

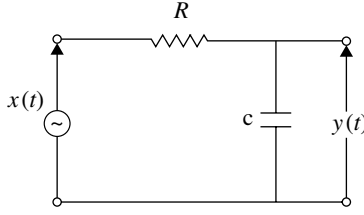
By performing the inverse Fourier transform of  $Y(\omega) = H(\omega)X(\omega)$ , we have the solution  $y(t)$ . Taking the Fourier transform of the both sides of Equation 2.95, and using a property of the Fourier transform (Eq. 2.33), an equation as same as Equation 2.96 is obtained. This is owing to facts that expressions of all variables as a form of  $A(\omega)e^{j\omega t}$  is equivalent to the Fourier transform and that the time function  $e^{j\omega t}$  is the eigen function of linear differential equations (including differentiation and integration).

A linear time-independent system described with a differential equation ( $b_0=1$ ) is expressed as a signal flow graph in Figure 2.22a. This graph (system) consists of two serially connected circuits. In the first circuit, input signal  $x(t)$  is differentiated with different orders and weighted summed (feed forward). The transfer function of this part becomes  $\sum_{m=0}^M a_m(j\omega)^m$ . In the second circuit, output signal  $y(t)$  is differentiated and weighted summed and added to the output of the first circuit (feedback). The transfer function of this part becomes  $b_0 / (1 - \sum_{n=1}^N b_n(j\omega)^n)$ . Because of the signal feedback, the impulse response spans in an infinite time period (IIR: infinite impulse response). If we have no feedback paths ( $b_1 = b_2 \cdots b_M = 0$ ), the impulse response remains in a finite time (FIR: finite impulse response). The FIR system is always stable since the input is merely differentiated and fed forward, while the IIR system is possibly unstable. The feedback system is decided to be stable or not by determining polarities of the poles of the transfer function (Section 2.3.2.3).

The IIR and FIR parts can be exchanged (Fig. 2.21b) without affecting the total characteristics, since a serial connection of linear circuits never depends on the order of connection. In this system, the differentiation circuits are commonly used.

Since the transfer function of a linear time-invariant system described with differential equation is easily obtained, to get the impulse response, it is a most convenient way for us to take the inverse Fourier transform of the transfer function as

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) e^{j\omega t} d\omega$$



**FIGURE 2.22** A first-order low-pass filter.

The integral is easily carried out by using the complex function theory, enjoying the fact that the transfer function  $H(\omega)$  is a rational function. When the transfer function is given with Equation 2.97, the impulse response is obtained as

$$h(t) = h_{FF}(t) * h_{FB}(t)$$

where we put  $h_{FF}(t) = \sum_{n=0}^N a_n \frac{d^n}{dt^n} \delta(t)$ ,  $h_{FB}(t) = \sum_{m=1}^M \alpha_m h_i(t)$ .

where  $h_i(t) = \begin{cases} e^{p_i t} & t > 0 \\ 0 & t < 0 \end{cases}$ ,  $p_i$  is the pole of  $H(s)$  ( $= H(j\omega)$ ), and  $\alpha_m$  is the coefficient when the denominator of  $H(\omega)$  is subjected to partial fraction decomposition. For a  $k$ -tuple root, we have (Eq. 2.93).

$$h_i(t) = \begin{cases} t^{k-1} e^{p_i t} / (k-1)! & t > 0 \\ 0 & t < 0 \end{cases}$$

### 2.3.4 Examples of Linear Systems

#### EXAMPLE 2.7 FIRST-ORDER LOW-PASS FILTER

We consider a circuit shown in Figure 2.22. The input  $x(t)$  is assumed to be a voltage source and the output voltage is denoted as  $y(t)$ . We have a linear first-order differential equation as

$$\frac{d}{dt} y(t) + ay(t) = ax(t) \quad \left( a = \frac{1}{RC} \right) \quad (2.98)$$

Solving the equation with an initial condition  $y(0) = v_0$ , we get

$$y(t) = a \int_0^t x(\tau) e^{-a(t-\tau)} d\tau + v_0 e^{-at}$$

To satisfy the condition of a linear system, we should set  $v_0 = 0$ .



Letting  $x(t) = \delta(t)$ , the impulse response is given as

$$h(t) = \begin{cases} ae^{-at} & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (2.99)$$

Taking the Fourier transform of the impulse response, we obtain the frequency transfer function as

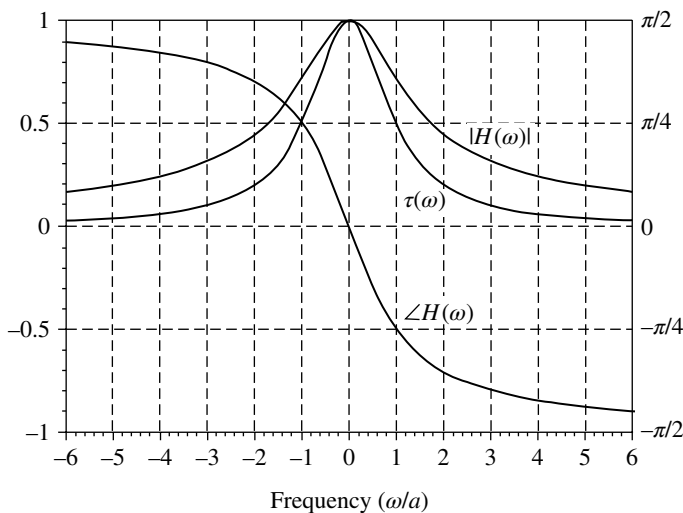
$$H(\omega) = \frac{1}{1 + j\omega/a}. \quad (2.100)$$

The amplitude  $|H(\omega)|$ , phase  $\angle H(\omega)$ , and group delay characteristics  $\tau(\omega)$  are shown in Figure 2.23. The frequency  $a$ , where the amplitude characteristics become  $|H(a)| = |H(0)|/\sqrt{2}$ , is called the ( $-3$  dB) cut-off frequency.

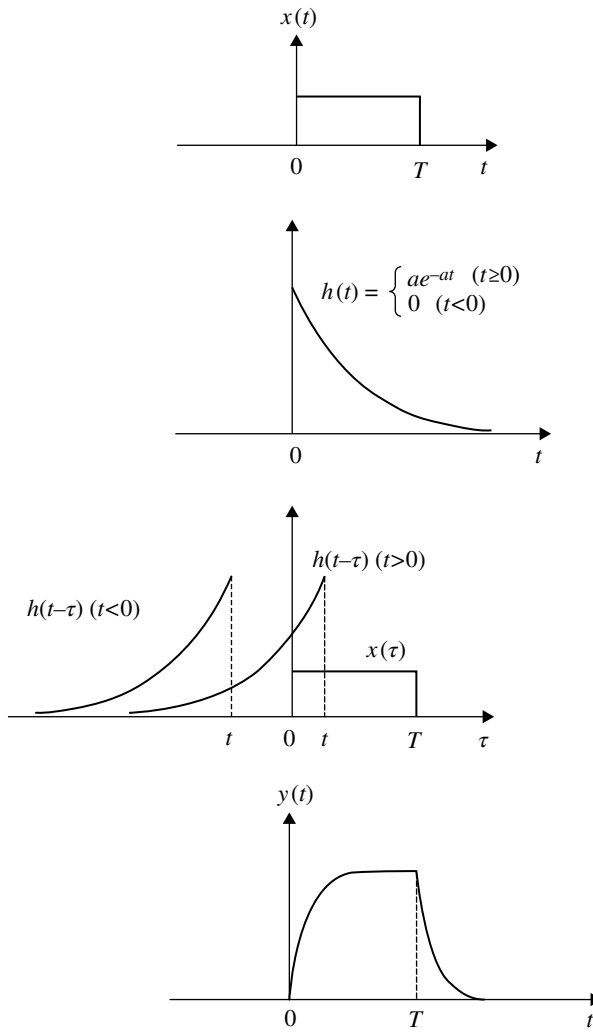
Assuming an input signal with a rectangular pulse as  $x(t) = \begin{cases} A & 0 \leq t \leq T \\ 0 & \text{otherwise} \end{cases}$ , we have the output signal as

$$y(t) = \begin{cases} 0 & t < 0 \\ A(1 - e^{-at}) & 0 \leq t \leq T \\ A(e^{-aT} - 1)e^{-at} & t > 0 \end{cases}$$

The output waveform is shown in Figure 2.24 with explanation of the convolution integral process.



**FIGURE 2.23** Frequency transfer function of a first-order low-pass filter.  $|H(\omega)|$  and  $\tau(\omega)$ : left scale,  $\angle H(\omega)$ : right scale.



**FIGURE 2.24** Explanation of a convolutional integral.

### EXAMPLE 2.8 DELAY LINE

A delay line with a time delay of  $t_0$  has the following impulse response

$$h(t) = \delta(t - t_0) \quad (2.101)$$

Applying input signal  $x(t)$ , the output signal becomes

$$\begin{aligned} y(t) &= x(t) * \delta(t - t_0) \\ &= x(t - t_0) \end{aligned}$$

The signal is delayed without distortion. Taking the Fourier transform of Equation 2.101, we have the transfer function as

$$H(\omega) = e^{-j\omega t_0} \quad (-\infty < \omega < \infty)$$

The transfer function shows the condition for a circuit to be free from distortion. A delay line or a distortion free circuit need not show the above condition in infinite frequency region but is sufficient to cover the frequency range of the input signal. The group delay defined with  $\tau(\omega) = -d\angle H(\omega)/d\omega$  becomes constant ( $\tau(\omega) = \tau_0$ ).

### EXAMPLE 2.9 INTEGRATE-AND-DUMP FILTER

The filter starts to integrate an input signal  $x(t)$  for a given time  $T$  and output the result. The output signal is written as

$$y(t) = \int_{t_0}^{t_0+T} x(\tau) d\tau \quad (t_0 \leq t \leq t_0 + T)$$

where  $t_0$  is an integration starting time. The starting and dumping times are controlled with external timing signals. If the integration is continuously done for a time span of  $T$ , we have

$$y(t) = \int_{t-T}^t x(\tau) d\tau \quad (-\infty < t < \infty)$$

This is the same as that for moving average in a time period  $T$  except for a constant  $T$ . The above equation is rewritten as

$$y(t') = \int_{-\infty}^{\infty} x(\tau) h(t' - \tau) d\tau$$

where  $h(t) = \begin{cases} 1 & 0 < t < T \\ 0 & \text{otherwise} \end{cases}$ .

Therefore, the moving average is performed by applying input signal to a filter with the above impulse response. The transfer function of the filter becomes

$$H(\omega) = \frac{T \sin(\omega T/2)}{\omega T/2} e^{-j\omega T/2} \quad (2.102)$$

where the term  $e^{-j\omega T/2}$  shows the time delay. The integrate-and-dump filter is used as the matched filter for a rectangular pulse signal (Section 3.3.3).

**EXAMPLE 2.10 THE HILBERT TRANSFORM CIRCUIT**

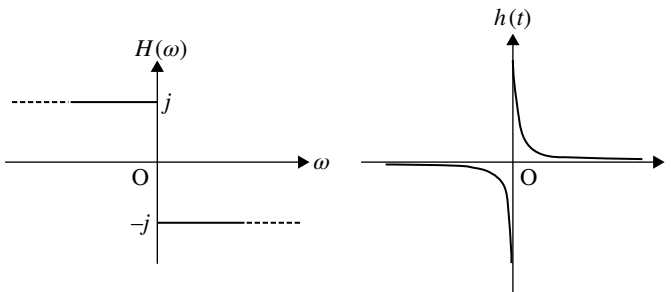
The circuit is a linear circuit with the transfer function  $H(\omega) = -j \operatorname{sgn}(\omega)$  (Fig. 2.25). Applying the symmetry property (Eq. 2.27) of the Fourier transform to Equation 2.44, the impulse response is given as

$$h(t) = \frac{1}{\pi t}$$

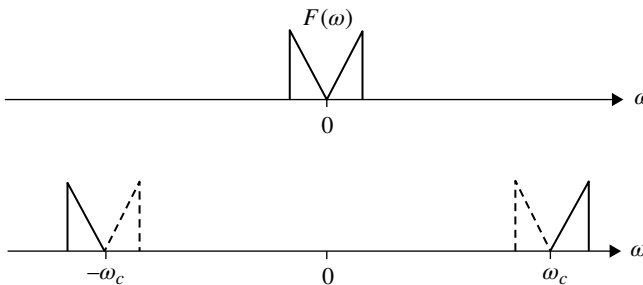
When the Hilbert transform is performed twice for a signal  $x(t)$ , the sign is reversed as  $-x(t)$ . Let us denote by  $\hat{f}(t)$  the Hilbert transform of  $f(t)$ . The following signal

$$s(t) = f(t) \cos \omega_c t \mp \hat{f}(t) \sin \omega_c t \quad (2.103)$$

is called an SSB signal. The spectrum  $S(\omega) (\leftrightarrow s(t))$  becomes a half side of the spectrum  $F(\omega) (\leftrightarrow f(t))$  and is shifted by the frequency  $\pm \omega_c$  (Fig. 2.26). The solid and broken lines correspond to the sign + and - in Equation 2.103, respectively. The Hilbert transform of a signal  $s(t) = f(t) \cos \omega_1 t$  becomes  $\hat{s}(t) = f(t) \sin \omega_1 t$ . The Hilbert transform of  $\hat{s}(t) = f(t) \sin \omega_1 t$  becomes  $-\hat{s}(t) = -f(t) \cos \omega_1 t$ , since it is obtained through taking the Hilbert transform twice on  $s(t) = f(t) \cos \omega_1 t$ . The SSB system is also used in an image rejection frequency converter (Section 8.3).



**FIGURE 2.25** Transfer function and impulse response of the Hilbert transform circuit.



**FIGURE 2.26** Spectrum of a single-side-band signal.

### EXAMPLE 2.11 FEEDBACK CIRCUIT

This system was invented by H. S. Black in 1927. The circuit block diagram is shown in Figure 2.27. The output signal is fed back to the input signal with the opposite polarity.

The transfer function becomes

$$H(\omega) = \frac{A(\omega)}{1 + A(\omega)\beta(\omega)}. \quad (2.104)$$

Although the feedback system has an inherent possibility of instability, it has some superior advantages as shown in the following.

*Stability* This topic is discussed in general in Section 2.3.2.3. Here, we find the stability condition on a simple assumption as

$$\beta(\omega) = \beta_0 \geq 0, \quad A(\omega) = \frac{A_0}{(1 + j\omega/a)} \quad (A_0, a \geq 0) \quad (2.105)$$

The transfer function  $A(\omega)$  is that for a first-order low-pass filter (Eq. 2.100). Putting  $s = j\omega$  and denoting by  $s_0$ , the pole or the root for the denominator of  $H(s)$ , we have

$$A(s_0)\beta_0(s_0) + 1 = 0$$

With the above equation and Equation 2.105, we get

$$s_0 = -a(A_0\beta_0 + 1) \leq 0$$

Thus, the system is always stable.

If we change the system as a positive feedback circuit, we get

$$H(\omega) = \frac{A(\omega)}{1 - A(\omega)\beta(\omega)}$$

This system is stable only if  $s_0 = a(A_0\beta_0 - 1) \leq 0$  or  $A_0\beta_0 \leq 1$  in contrast to the negative feedback system. As in a radio wave repeater (Section 9.1.6) where a part of the output signal is fed back to the input, and the phase of the fed-back signal is not controllable, the stability condition becomes strict because of the possibility of positive feedback.

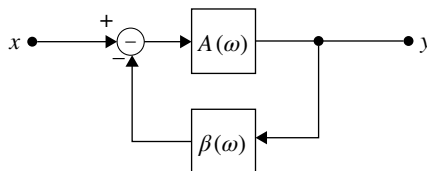
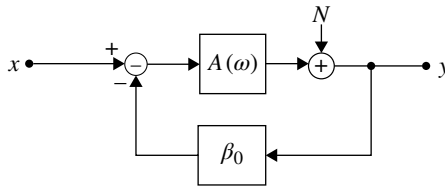


FIGURE 2.27 A feedback circuit.



**FIGURE 2.28** Negative feedback circuit with noise.

*Improvement of frequency transfer characteristics* When  $|A(\omega)\beta(\omega)| \gg 1$ , we have  $H(\omega) \approx 1/\beta(\omega)$  and if  $\beta(\omega) = \beta_0$ , the transfer function becomes  $H(\omega) \approx 1/\beta_0$ . Thus we can get approximately a flat frequency response regardless of the main path transfer function  $A(\omega)$  (equalization of linear distortion). We should note that the system gain is reduced to  $1/\beta_0$ .

*Noise suppression* We consider a system where noise is added as shown in Figure 2.28. Letting  $x(t) \leftrightarrow X(\omega)$ ,  $y(t) \leftrightarrow Y(\omega)$ , we have

$$Y(\omega) = \frac{A(\omega)}{1 + A(\omega)\beta_0} X(\omega) + \frac{N(\omega)}{1 + A(\omega)\beta_0}$$

We understand that the noise level is reduced by a factor of  $1/(1 + A(\omega)\beta_0)$  owing to the negative feedback. Nonlinear distortion instead of noise can be reduced as shown in Section 8.6.3.

*Inverse circuit* A negative feedback circuit becomes the inverse system of a linear or nonlinear feed forward system equalizing its transfer characteristics as described in the following example.

## EXAMPLE 2.12 INVERSE CIRCUIT

Input signal is distorted if it is applied to a circuit ( $H(\omega)$ ) that never satisfies the distortion free condition (Example 2.8). The inverse circuit ( $H^{-1}(\omega)$ ) is serially connected to the circuit to fulfill the distortion-free condition as a whole. The overall transfer function becomes  $H(\omega)H^{-1}(\omega) = Ae^{-j\omega t_0}$  ( $A, t_0$ : const.). The inverse circuit for a given system with  $H(\omega) = 1 + F(\omega)$  is shown in Figure 2.29 for an example. The inverse circuit becomes unstable depending on some conditions, since it is a feedback system. For example, let us assume for the given circuit, an impulse response  $h(t) = 1 + b\delta(t - \tau)$ , that is,  $F(\omega) = be^{-j\omega\tau}$ , then for the inverse circuit, we have the following transfer function:

$$\begin{aligned} H^{-1}(\omega) &= \frac{1}{1 + be^{-j\omega\tau}} \\ &= \sum_{n=0}^{\infty} (-b)^n e^{-jn\omega\tau} \end{aligned}$$

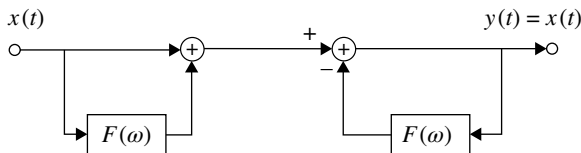
The impulse response of the inverse circuit becomes  $h^{-1}(t) = \sum_{n=0}^{\infty} (-b)^n \delta(t - n\tau)$  ( $\leftrightarrow H^{-1}(\omega)$ ). If a bounded input signal  $z(t)$  is applied to the circuit, the output signal is  $y(t) = \sum_{n=0}^{\infty} b^n z(t - n\tau)$ . Then we have

$$|y(t)| \leq \sum_{n=0}^{\infty} |b|^n |z(t - n\tau)| \leq z_M \sum_{n=0}^{\infty} |b|^n \quad (z_M = |z(t)|_{\max})$$

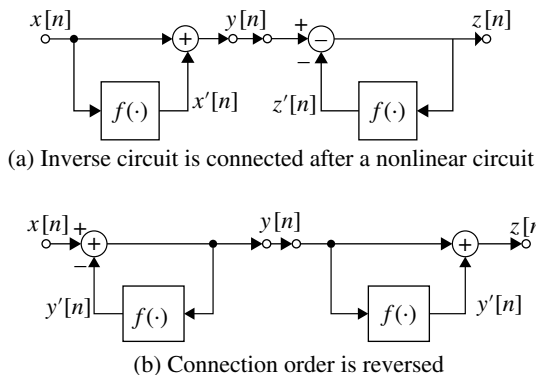
Thus, the inverse circuit is stable if  $|b| < 1$ .

Since  $H(\omega) \leftrightarrow h(t) = 1 + b\delta(t - \tau)$ , the output signal of the circuit with  $H(\omega)$  becomes  $z(t) = x(t) + bx(t - \tau)$ . This system is a model where broadcasted TV signals, for example, reach a receiver directly and in a time delay (echo) due to reflection from an obstacle like a mountain. If the TV signal is analog and amplitude modulated, the receiver shows the same two pictures (called a ghost), which is separated in the horizontal direction. The ghost signal is removed with the inverse circuit: we can confirm that  $y(t) = x(t)$  in the time region. In this example, the given circuit and the inverse circuit are called a feed forward and feedback circuits, respectively. The order of connection of these circuit never affects the overall characteristics, since both of them are a linear circuit.

Next, we consider a system that includes a feed forward nonlinear circuit  $f^*$  as shown in Figure 2.30a. Here, we assume a time-discrete system for simplicity. The nonlinear circuit is assumed to have a memory for a time length of  $M$ , that is, the



**FIGURE 2.29** An example of the inverse circuit connection.



**FIGURE 2.30** Connection of inverse circuit including nonlinear circuit.





circuit, we can confirm that the output  $z$  of the inverse circuit becomes the input  $x$  to the given circuit as  $z = f^{-1}(f(x)) = f^{-1}(y) = x$ . Next, we connect to the inverse circuit in front of the given circuit. Applying an input  $y$  to the inverse circuit, we have the output  $z = f(f^{-1}(y)) = f(x) = y$ . Thus, the order of connection of the inverse circuit never affects the overall characteristics even for a nonlinear circuit. This property is used for identification (learning) of a given nonlinear circuit (Section 8.6.3).

## 2.4 DISCRETE-TIME SYSTEM

Extracting signal from a continuous time signal at periodic time instants is called *sampling*. The sampled signal is called *discrete-time signal*. Assigning a finite discrete level to the sampled signal is called *quantization*. Expressing the quantized signal with a code  $\{1,0\}$  is called coding. Thus, the digitization of an analog signal is completed. Digitized signal has various advantages over analog signal in treatments such as signal processing, storage/retrieval, and transmission (communications).

The discrete-time signal is considered to be a special case of analog signal in a sense that it is produced from the analog signal through sampling. Therefore, all the properties of time-continuous signals and systems hold equally for discrete-time signals. The difference appears in mathematical expression: The expression becomes simple in a sense that differentiation and integrals in the analog signal expression disappear and only multiplication and summation are used for the discrete-time signal expression. Here, we discuss the properties of discrete-time signals and systems.

### 2.4.1 Sampling and the Sampling Theorem

Consider an continuous-time signal  $x(t)$ . From the signal, we extract discrete-time signal  $x(nT)$  at a time interval of  $T$  (sampling). We denote as  $x[n]$  if the sampling period  $T$  is not explicitly shown. Signal  $x[n]$  is a data sequence and ceases to be a function of time  $t$ . When we express a discrete-time signal as a function of time  $t$ , we use the sampling function  $f_s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$  that is a series of the delta function as

$$\begin{aligned} x_s(t) &= x(t)f_s(t) \\ &= x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT) \end{aligned} \quad (2.106)$$

$$= \sum_{n=-\infty}^{\infty} x(nT)\delta(t - nT) \quad (2.107)$$

where the property of the delta function (Eq. 2.9) is used. The delta function never expresses the actually sampled signal, since the delta function takes an infinite value. It is approximated with a rectangular pulse signal with a finite height and small width. When a mathematical limit is taken, the delta function should be used: the

integral of a sampled signal becomes zero if the width of the pulse signal approaches zero while taking a finite height, and the analog signal is never recovered from the discrete time signal.

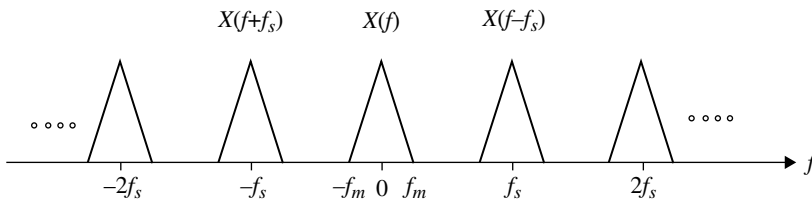
When the sampling is done, the signals are discarded except for the signal at the sampling instants. The sampling theorem assures that the original continuous-time signal can be restored from the sampled signal. The theorem states that if a time-continuous signal is bandlimited below  $f_m$  (Hz), the sampling frequency should be higher than  $2f_m$  (Hz). This is proved as follows.

The Fourier transform of Equation 2.106 becomes as follows from Equations 2.18 and 2.40

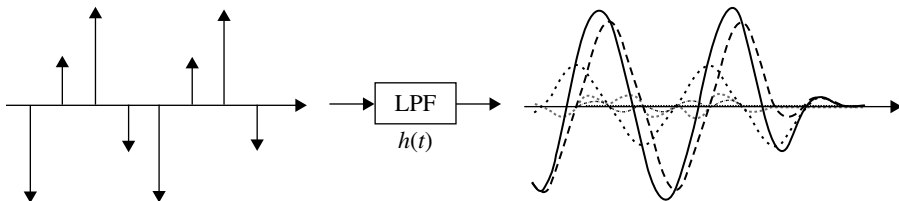
$$\begin{aligned} X_s(\omega) &= \frac{1}{2\pi} X(\omega) * F_s(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega - x) \omega_s \sum_{n=-\infty}^{\infty} \delta(x - n\omega_s) dx \\ &= \frac{1}{T} \sum_{n=-\infty}^{\infty} X(\omega - n\omega_s) \end{aligned} \quad (2.108)$$

where  $\omega_s = 2\pi/T$ ,  $x_s(t) \leftrightarrow X_s(\omega)$ , and  $f_s(t) \leftrightarrow F_s(\omega)$ .

The spectrum  $X_s(\omega)$  consists of  $X(\omega)$  repeated with a spacing of the sampling frequency  $1/T (= f_s)$  (Fig. 2.32). If the spectra  $X(\omega - n\omega_s)$  produced by the sampling never overlap with each other, it is intuitively understood, from Figure 2.32, that the original time-continuous signal is recovered by applying the sampled signal into a low-pass (interpolation) filter that passes only the fundamental frequency component  $X(\omega)$  (Fig. 2.33). The sampling theorem states the condition that the sampled signal spectra should not overlap. If the spectra overlap, the recovered signal differs from the original one including an error signal called aliasing error. To prevent the aliasing error, the input signal should be bandlimited through a (anti-aliasing) filter before sampling.



**FIGURE 2.32** The spectra of a sampled signal.



**FIGURE 2.33** Restoration of the original time-continuous signal.

The low-pass (interpolation) filter output signal becomes

$$y(t) = x_s(t) * h_{\text{LPF}}(t)$$

From Equation 2.107, we have

$$y(t) = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(nT) \delta(\tau - nT) h_{\text{LPF}}(t - \tau) d\tau = \sum_{n=-\infty}^{\infty} x(nT) h_{\text{LPF}}(t - nT)$$

where  $h_{\text{LPF}}(t)$  is the impulse response of the low-pass filter. It can be arbitrarily chosen as long as it passes the fundamental component without distortion and suppresses the higher order components. Since the coefficient of  $1/T$  appears in the sampling process, a gain of  $T$  should be multiplied to get  $y(t) = x(t)$ . When the sampling frequency is the minimum, that is,  $2f_m$ , called the Nyquist (sampling) frequency, the low-pass filter should take the ideal rectangular frequency transfer characteristics that pass frequency components  $-f_m \sim f_m$ . Then the output signal becomes

$$y(t) = \sum_{n=-\infty}^{\infty} x(nT) \frac{\sin 2\pi f_m(t - nT)}{2\pi f_m(t - nT)}$$

where  $T$  is  $1/(2f_m)$  and  $x(nT)$  are the sampled signals. The ideal filter is not actually realizable and some error appears. As the sampling frequency increases, the realization of the low-pass filter becomes easier owing to relaxed requirements for the filter characteristics.

If a complex signal, for example,  $\exp(j\omega_0 t)$  ( $\omega_0 > 0$ ) is considered, its spectrum appease only in positive frequency region. In this case, the spectra never overlap and therefore the sampling theorem is fulfilled even for  $\omega_s / 2 \leq \omega_0 < \omega_s$  ( $\omega_s$ : sampling frequency). The actual signal is real and its spectrum appears also in the negative frequency region as a form of complex symmetry (Eq. 2.37). Thus, to realize the circuit, the complex signal should be upsampled (Section 2.4.7) to satisfy the sampling theorem. The complex signal expression is used for convenience of mathematical notation in signal processing such as in single-sideband frequency conversion (Section 8.3).

**The Frequency Sampling Theorem** The Fourier transform (spectrum) of a periodic signal (function) is given by the delta function series (line spectra) in the frequency domain (Section 2.1.2). Here, we show that the spectrum  $F_T(\omega)$  of a signal  $f_T(t)$  produced by extracting from a periodic signal for a periodic time interval and setting to null at the other time can be expressed with the line spectra  $F_T(n\omega_0)$  ( $\omega_0 = 2\pi/T, n = -\infty \sim \infty$ ) of the periodic function. The function  $f_T(t)$  is written as

$$f_T(t) = h(t) \sum_{n=-\infty}^{\infty} f_T(t - nT)$$

where

$$h(t) = \begin{cases} 1 & |t| \leq T/2 \\ 0 & \text{otherwise} \end{cases}$$

Taking the Fourier transform of the above equation, we have the target result as

$$\begin{aligned} F_T(\omega) &= \frac{1}{2\pi} H(\omega) * \omega_0 \sum_{n=-\infty}^{\infty} F_T(n\omega_0) \delta(\omega - n\omega_0) \\ &= \frac{1}{T} \sum_{n=-\infty}^{\infty} F_T(n\omega_0) H(\omega - n\omega_0) \\ &= \sum_{n=-\infty}^{\infty} F_T(n\omega_0) \frac{\sin(\omega - n\omega_0)T/2}{(\omega - n\omega_0)T/2} \end{aligned}$$

where  $h(t) \leftrightarrow H(\omega) = \frac{T \sin \omega T/2}{\omega T/2}$

We consider a periodic function of  $f_{T'}(t)$  generated by expanding the time region of  $f_T(t)$  from  $-T/2 \leq t \leq T/2$  to  $-T'/2 \leq t \leq T'/2$  ( $T' > T$ ) and setting null at the expanded region. Then, we can develop the same arguments as in the above discussion. The extrapolating function  $h(t)$  can be arbitrary as long as it takes 1 for  $-T/2 \leq t \leq T/2$ .

The above property is called the frequency sampling theorem, since it has a symmetrical relation with the sampling theorem in time domain.

## 2.4.2 The Energy, Power, and Correlation of Discrete-Time Signals

We consider a time-discrete signal  $x[n]$  ( $n = 0, 1, \dots, N$ ). The energy  $E$ , average power  $P$ , and auto-correlation function  $R_{xx}[m]$  are given as

$$\begin{aligned} E &= \sum_{n=0}^N x^2[n] \\ P &= \frac{1}{N+1} \sum_{n=0}^N x^2[n] \\ R_{xx}[m] &= \sum_{n=0}^N x[n+m]x[n] \end{aligned}$$

where we put  $x[n+m] = 0$  for  $n+m < 0$ ,  $n+m > N$ .

When  $x[n]$  taking nonzero values continues infinitely, the energy becomes infinite. Then, the average power and the autocorrelation function are defined as

$$\begin{aligned} P &= \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x^2[n] \\ R_{xx}[m] &= \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x[n+m]x[n] \end{aligned}$$

Consider another discrete-time signal  $y[n]$ . Then the cross-correlation function  $R_{xy}[m]$  is defined as

$$R_{xy}[m] = \sum_{n=-\infty}^{\infty} x[n+m]y[n]$$

When  $x[n]$  and  $y[n]$  continue infinitely, we use

$$R_{xy}[m] = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x[n+m]y[n]$$

The inner product in a time period of  $n=N_1 \sim N_2$  is defined as

$$R_{xy}[0] = \sum_{n=N_1}^{N_2} x[n]y[n]$$

When the inner product becomes zero, we call that the two signals are orthogonal with each other in the time region of  $n=N_1 \sim N_2$ . For an example, any pair of the (sampled) Walsh functions (Section 2.1.3) is orthogonal.

### 2.4.3 The Fourier Transform of Discrete-Time Signals

**The discrete-time Fourier transform:** We take the Fourier transform of a discrete-time signal  $x_s(t)$  which is generated by sampling an continuous-time signal  $x_s(t)$ . The result is already shown with Equation 2.108. Here, we derive another expression. Performing the Fourier integral of Equation 2.107, we have the discrete-time Fourier transform (DTFT) formula as

$$X_s(\omega) = \sum_{n=-\infty}^{\infty} x(nT)e^{-jn\omega T} \quad (2.109)$$

This is a periodic function of  $\omega$  with a period of  $2\pi/T$ . Taking the inverse Fourier transform of the above equation, we have  $x_s(t)$  as

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} X_s(\omega) e^{j\omega t} d\omega &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(nT) e^{-jn\omega T} e^{j\omega t} d\omega \\ &= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} x(nT) \int_{-\infty}^{\infty} e^{j(t-nT)\omega} d\omega \\ &= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} x(nT) \lim_{\Omega \rightarrow \infty} \int_{-\Omega}^{\Omega} e^{j(t-nT)\omega} d\omega \\ &= \sum_{n=-\infty}^{\infty} x(nT) \lim_{\Omega \rightarrow \infty} \frac{\sin \Omega(t-nT)}{\pi(t-nT)} \\ &= \sum_{n=-\infty}^{\infty} x(nT) \delta(t-nT) = x_s(t) \quad (\text{from Eq. 2.5}) \end{aligned}$$

Next, we express  $x(nT) (\equiv x[n])$  with  $X_s(\omega)$ . We rewrite Equation 2.109 as

$$X_s(\Omega) = \sum_{n=-\infty}^{\infty} x[n] e^{-jn\Omega} \quad (\text{DTFT}) \quad (2.110)$$

where we put  $\Omega = \omega T$ . This equation is a periodic function of  $\Omega$  with a period of  $2\pi$  and can be thought as the Fourier series expansion with coefficients of  $x[n]$ . Multiplying  $e^{jm\Omega}$  to the both sides of Equation 2.110, using the orthogonality,

$$\int_{-\pi}^{\pi} e^{j(m-n)\Omega} d\Omega = \begin{cases} 2\pi & m = n \\ 0 & m \neq n \end{cases} \quad (2.111)$$

and integrating with respect to  $\Omega$  in a range of  $-\pi \sim \pi$ , we have

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X_s(\Omega) e^{jn\Omega} d\Omega \quad (\text{inverse DTFT})$$

We use the notation  $x[n] \leftrightarrow X_s(\Omega)$  (or  $X_s(\omega)$ ) for the relation between  $x[n]$  and  $X_s(\Omega)$  (or  $X_s(\omega)$ ) in the following.

**The z-transform:** Letting  $x(t)=0$  for  $t < 0$  and taking the Laplace transform of Equation 2.107, we have

$$\begin{aligned} X_s(s) &= \int_0^{\infty} \sum_{n=0}^{\infty} x(nT) \delta(t-nT) e^{-st} dt \\ &= \sum_{n=0}^{\infty} x[n] e^{-snT} \end{aligned} \quad (2.112)$$

where we put  $s = \sigma + j\omega$ . If we take the inverse Laplace transform of Equation 2.112, we have  $x_s(t) = \sum_{n=0}^{\infty} x[nT] \delta(t-nT)$ .

If we put  $z = e^{sT} = e^{(\sigma + j\omega)T}$  in Equation 2.112, we have the following expression:

$$X(z) = \sum_{n=0}^{\infty} x[n] z^{-n} \quad (z\text{-transform}) \quad (2.113)$$

Getting  $X(z)$  from  $x[n]$  is called the inverse z-transform. We rewrite Equation 2.113 as

$$X_s(s) = \sum_{n=0}^{\infty} x[n] e^{-\sigma nT} e^{-jn\omega T}$$

Putting  $\omega T = \Omega$ , multiplying  $e^{jm\Omega}$  to the both sides of the above equation, integrating in a range of  $-\pi \sim \pi$  and using the orthogonality (Eq. 2.111), we have

$$\begin{aligned} \int_{-\pi}^{\pi} X(\Omega) e^{jm\Omega} d\Omega &= \sum_{n=0}^{\infty} x[n] e^{-\sigma nT} \int_{-\pi}^{\pi} e^{j(m-n)\Omega} d\Omega \\ &= 2\pi x[m] e^{-\sigma mT} \end{aligned}$$

From this, we get

$$x[m] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\Omega) e^{m(\sigma T + j\Omega)} d\Omega$$

Using  $s = \sigma T + j\Omega$ , we have

$$x[m] = \frac{1}{2\pi j} \int_{\sigma T - j\pi}^{\sigma T + j\pi} X(s) e^{ms} ds$$

This equation is called the inverse discrete-time Laplace transform. We use  $z$  instead of  $\Omega$ . Then, since  $dz = jz d\Omega$  from  $z = e^{\sigma T + j\Omega}$ , we have

$$x[m] = \frac{1}{2\pi j} \oint X(z) z^{m-1} dz \quad (\text{inverse } z\text{-transform})$$

where the contour integral is carried out on the path  $z = e^{\sigma T + j\Omega} (\Omega: -\pi \rightarrow \pi)$  (anti-clockwise). This result can be obtained also by applying the following formula of complex integral to Equation 2.113.

$$\frac{1}{2\pi j} \oint z^{n-1} dz = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$$

The parameter  $\sigma$  that is introduced to secure convergence in the  $z$ -transform never appears explicitly. If we put  $\sigma = 0$ , the  $z$ -transform becomes the DTFT. The Fourier transform gives a more direct description of a system, for example, the frequency transfer function than the Laplace transform. In this book, we never focus on the difference between them, and the notation  $z$  means either  $z = e^{(\sigma + j\omega)T}$  or  $z = e^{j\omega T}$ . The relation between  $x[n]$  and  $X(z)$  is denoted as  $x[n] \leftrightarrow X(z)$ .

**2.4.3.1 Properties of the Discrete-Time Fourier Transform** The expressions in the time-continuous Fourier transform are changed as in the following.

- (1) Linearity  $x[n] = a_1 x_1[n] + a_2 x_2[n] \leftrightarrow X(\Omega) = a_1 X_1(\Omega) + a_2 X_2(\Omega)$   
where  $x_1[n] \leftrightarrow X_1(\Omega)$ ,  $x_2[n] \leftrightarrow X_2(\Omega)$
- (2) Time shift

$$x[n-m] \leftrightarrow X(\Omega) e^{-jm\Omega} \quad (2.114)$$

$$(\text{proof}) \quad \sum_{n=-\infty}^{\infty} x[n-m] e^{-jn\Omega} = \sum_{k=-\infty}^{\infty} x[k] e^{-jk\Omega} \cdot e^{-jm\Omega} = X(\Omega) e^{-jm\Omega}$$

- (3) Convolution

$$x[n] = \sum_{m=-\infty}^{\infty} x_1[n-m] x_2[m] \leftrightarrow X_1(\Omega) X_2(\Omega) \quad (2.115)$$

(proof)

$$\begin{aligned}
\sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} x_1[n-m]x_2[m]e^{-jn\Omega} &= \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} x_1[k]e^{-jk\Omega}x_2[m]e^{-jm\Omega} \\
&= \sum_{k=-\infty}^{\infty} x_1[k]e^{-jk\Omega} \sum_{m=-\infty}^{\infty} x_2[m]e^{-jm\Omega} = X_1(\Omega)X_2(\Omega)
\end{aligned}$$

(4) Multiplication

$$x_1[n]x_2[n] \leftrightarrow \frac{1}{2\pi} X_1(\Omega) * X_2(\Omega) \quad (2.116)$$

$$\begin{aligned}
X(\Omega) &= \sum_{n=-\infty}^{\infty} x_1[n]x_2[n]e^{-jn\Omega} \\
(\text{proof}) \quad &= \sum_{n=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} X_1(x)e^{jnx}dx \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} X_2(y)e^{jny}dy \cdot e^{-jn\Omega} \\
&= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} X_1(x)X_2(y) \sum_{n=-\infty}^{\infty} e^{jn(x+y-\Omega)} dx dy
\end{aligned}$$

From Equation 2.19, we have

$$\sum_{n=-\infty}^{\infty} e^{jn(x+y-\Omega)} = 2\pi\delta(x+y-\Omega) \quad (|x+y-\Omega| \leq \pi)$$

$$\text{Thus, we get } X(\Omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X_1(x)X_2(\Omega-x)dx = \frac{1}{2\pi} X_1(\Omega) * X_2(\Omega)$$

(5) Frequency shift

$$x_1[n]e^{jn\Delta\Omega} \leftrightarrow X_1(\Omega - \Delta\Omega) \quad (2.117)$$

$$(\text{proof}) \quad \sum_{n=-\infty}^{\infty} x_1[n]e^{jn\Delta\Omega}e^{-jn\Omega} = \sum_{n=-\infty}^{\infty} x_1[n]e^{-jn(\Omega-\Delta\Omega)} = X_1(\Omega - \Delta\Omega)$$

**Properties of the z-transform:** Since the z-transform is given from the discrete-time Fourier integral assuming  $x[n] = 0 (n < 0)$ ,  $x[\infty] = 0$ , and letting  $e^{j\Omega} = z$  formally, the properties of the DTFT hold equally for the z-transform for signals  $x[\infty] = 0$ .

### 2.4.3.2 Examples of the Discrete-Time Fourier Integrals and z-Transform

(1) Rectangular signal

$$\begin{aligned}
x[n] &= \begin{cases} 1 & 0 \leq n \leq N-1 \\ 0 & n < 0 \end{cases} \\
X(z) &= \sum_{n=0}^{N-1} z^{-n} = \frac{1-z^{-N}}{1-z^{-1}}
\end{aligned}$$



(2) Direct current signal  $x[n] = 1$  ( $-\infty \leq n \leq \infty$ )

$$X(\omega) = \lim_{N \rightarrow \infty} \sum_{n=-\infty}^{\infty} e^{-jn\omega T} = \frac{2\pi}{T} \sum_{n=-\infty}^{\infty} \delta\left(\omega - \frac{2n\pi}{T}\right) \quad (-\infty \leq \omega \leq \infty) \quad (\text{from Eq. 2.19}) \quad (2.118)$$

(3) The signum function

$$\text{sgn}[n] = \begin{cases} 1 & n > 0 \\ 0 & n = 0 \\ -1 & n < 0 \end{cases}$$

The DTFT becomes

$$\text{SGN}(\omega) = - \sum_{n=-\infty}^{-1} e^{-jn\omega T} + \sum_{n=1}^{\infty} e^{-jn\omega T}$$

Letting  $z = e^{j\omega T}$ , we have

$$\begin{aligned} \text{SGN}(\omega) &= \sum_{n=0}^{\infty} (z^{-n} - z^n) = \lim_{N \rightarrow \infty} \left( \frac{1 - z^{-(N+1)}}{1 - z^{-1}} - \frac{1 - z^{(N+1)}}{1 - z} \right) \\ &= \frac{1}{1 - z^{-1}} - \frac{1}{1 - z} + \lim_{N \rightarrow \infty} \left( \frac{-z^{-(N+1)}}{1 - z^{-1}} + \frac{z^{(N+1)}}{1 - z} \right) \\ &= \frac{1 + z^{-1}}{1 - z^{-1}} + \lim_{N \rightarrow \infty} \left( \frac{-z^{-(N+1/2)}}{(z^{1/2} - z^{-1/2})} + \frac{z^{(N+1/2)}}{z^{-1/2} - z^{1/2}} \right) \\ &= \frac{1 + z^{-1}}{1 - z^{-1}} + \lim_{N \rightarrow \infty} \left( \frac{-e^{-j(N+1/2)\omega T}}{2j \sin(\omega T/2)} + \frac{e^{j(N+1/2)\omega T}}{-2j \sin(\omega T/2)} \right) \\ &= \frac{2 \cos(\omega T/2)}{2j \sin(\omega T/2)} - \lim_{N \rightarrow \infty} \frac{\cos(N+1/2)\omega T}{j \sin(\omega T/2)} \\ &= \frac{1}{j \tan(\omega T/2)} \end{aligned} \quad (2.119)$$

where we used  $\lim_{N \rightarrow \infty} \cos(N\omega T) = 0$  (Eq. 2.B.3). The above result looks different from the following one given with Equations 2.44 and 2.108.

$$\text{SGN}(\omega) = \frac{1}{T} \sum_{n=-\infty}^{\infty} \frac{2}{j(\omega - n\omega_0)} \quad \left( \omega_0 = \frac{2\pi}{T} \right)$$

However, we can confirm that the results are the same, if we use the following mathematical formula:

$$\lim_{N \rightarrow \infty} \sum_{n=-N}^N \frac{1}{x + n} = \frac{\pi}{\tan \pi x}$$

(4) The unit step function

$$u[n] = \begin{cases} 1 & n \geq 0 \\ 0 & n < 0 \end{cases}$$

We put  $N \rightarrow \infty$  in the rectangular pulse signal. Since  $\lim_{N \rightarrow \infty} z^{-N} = 0$ , the  $z$ -transform becomes

$$U(z) = \sum_{n=0}^{\infty} z^{-n} = \frac{1}{1-z^{-1}} \quad (|z| < 1)$$

Taking  $\omega$  for the variable, the DTFT is given as follows:

$$U(\omega) = \lim_{N \rightarrow \infty} \left. \frac{1-z^{-N}}{1-z^{-1}} \right|_{z=e^{j\omega T}} = \lim_{N \rightarrow \infty} \left. \frac{1}{1-z^{-1}} - \frac{z^{-N}}{1-z^{-1}} \right|_{z=e^{j\omega T}}$$

where  $\left. \frac{z^{-N}}{1-z^{-1}} \right|_{z=e^{j\omega T}}$  is a periodic function of  $\omega$  with a period of  $2\pi/T$ . In a range of  $|\omega| \leq \pi/T$ , we get

$$\begin{aligned} \lim_{N \rightarrow \infty} \left. \frac{z^{-N}}{1-z^{-1}} \right|_{z=e^{j\omega T}} &= \lim_{N \rightarrow \infty} \frac{e^{-jN\omega T}}{1-e^{-j\omega T}} = \lim_{N \rightarrow \infty} \frac{e^{-j(N-1/2)\omega T}}{2j \sin(\omega T/2)} \\ &= \lim_{N \rightarrow \infty} \frac{\omega T}{2 \sin(\omega T/2)} \frac{e^{-j(N-1/2)\omega T}}{j\omega T} = \frac{\omega T}{2 \sin(\omega T/2)} (-\pi \delta(\omega T)) \\ &= -\pi \delta(\omega T) = -\frac{\pi}{T} \delta(\omega) \quad (\text{from Eq. 2.5}) \end{aligned}$$

Thus, we have

$$U(\omega) = \frac{1}{1-e^{-j\omega T}} + \frac{\pi}{T} \sum_{n=-\infty}^{\infty} \delta\left(\omega - n \frac{2\pi}{T}\right)$$

This result is different from the one given by applying the time-continuous Fourier integral (Eq. 2.46) to Equation 2.108. To equate the results, the definition should be modified as

$$u'(n) = \frac{1}{2} + \frac{1}{2} \text{sgn}[n] = \begin{cases} 1 & n > 0 \\ 1/2 & n = 0 \\ 0 & n < 0 \end{cases}$$

With this result, Equations 2.118 and 2.119, we get

$$U'(\omega) = \frac{1}{j2 \tan(\omega T/2)} + \frac{\pi}{T} \sum_{n=-\infty}^{\infty} \delta\left(\omega - n \frac{2\pi}{T}\right)$$

(5) Exponential function

$$x[n] = \begin{cases} e^{-anT} & (n \geq 0) \\ 0 & (n < 0) \end{cases} \quad (2.120)$$

The time-discrete Fourier transform becomes

$$X(z) = \sum_{n=0}^{\infty} e^{-anT} z^{-n} = \frac{1}{1 - e^{-aT} z^{-1}} \quad (a \leq 0) \quad (2.121)$$

The  $z$ -transform becomes

$$X(z) = \sum_{n=0}^{\infty} e^{-anT} z^{-n} = \frac{1}{1 - e^{-aT} z^{-1}} \quad (\text{Re}(z) \geq |a|) \quad (2.122)$$

#### 2.4.4 Response of Discrete-Time System

We discuss response of systems sampled at time instants  $t = nT$  ( $-\infty < n < \infty$ ) (Fig. 2.34).

An input signal  $x(t)$  is sampled as

$$x_s(t) = x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT)$$

From the property of the delta function, the above result is rewritten as

$$x_s(t) = \sum_{n=-\infty}^{\infty} x(nT) \delta(t - nT)$$

Applying  $x_s(t)$  to a linear time-invariant system with impulse response of  $h(t)$ , we get the output  $y(t)$  as

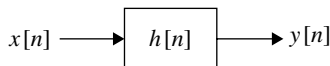
$$y(t) = \sum_{n=-\infty}^{\infty} x(nT) h(t - nT)$$

Sampling  $y(t)$  at  $t = mT$ , we have

$$y(mT) = \sum_{n=-\infty}^{\infty} x(nT) h(mT - nT)$$

or

$$y(mT) = \sum_{n=-\infty}^{\infty} x(mT - nT) h(nT)$$



**FIGURE 2.34** Response of time-discrete system.

If the causality holds, we get

$$y(mT) = \sum_{n=0}^{\infty} x(mT - nT)h(nT)$$

If the sampling period  $T$  is omitted, the expression becomes

$$\begin{aligned} y[m] &= \sum_{n=-\infty}^{\infty} x[n]h[m-n] \\ &= \sum_{n=-\infty}^{\infty} x[m-n]h[n] = x[m] * h[m] \end{aligned} \quad (2.123)$$

This operation corresponds to the convolutional integral (Eqs. 2.90 or 2.91) in a time-continuous system and is called convolution. If we let  $x[n] = \delta[n]$ , we have the impulse response  $y[m] = h[m]$ , where  $\delta[n]$  takes 1 for  $n=0$  and 0 otherwise (the Kronecker delta).

**Linearity:** For a sampled (time-discrete) system, the linearity of the system is stated as follows: We assume that the output signals  $y_1[n]$  and  $y_2[n]$  for input signals are  $x_1[n]$  and  $x_2[n]$ , respectively. Then the linearity holds if for a linearly combined input signal  $x[n] = a_1x_1[n] + a_2x_2[n]$  ( $a_1, a_2$ : constant), the output signal becomes  $y[n] = a_1y_1[n] + a_2y_2[n]$ . The number of linearly combined signals can be expanded to an arbitrary one. The linearity requires at least that the output signal should be zero when the input signal is zero

**Time-invariant system:** The time-invariant system is defined as follows: consider a system where the input signal is  $x[n]$  and output signal is  $y[n]$ . Then for a time-delayed input signal  $x[n-N]$ , the output signal also delayed as  $y[n-N]$ .

From the linearity and time-invariant properties of a time-discrete system, the input and output relations given with Equation 2.123 can be also given as follows. We denote the output signal for an input signal of  $\delta[n]$ , that is, the impulse response of the linear time-invariant system by  $h[n]$ . An input signal is expressed with a linear combination as  $x[n] = \sum_{m=-\infty}^{\infty} x[m]\delta[n-m]$ . When the signal is applied to the system, the output signal becomes  $y[n] = \sum_{m=-\infty}^{\infty} x[m]h[n-m]$ .

**Causality:** A system is called causal if for an input signal  $x[m] = 0$  [ $m < M$ ], the output signal becomes  $y[m] = 0$  [ $m < M$ ]. This means that the output (result) never appears before an input (cause). Applying this condition to Equation 2.123, we have the condition for the impulse response as follows:

$$h[m] = 0 \quad (m < 0).$$

### EXAMPLE 2.13 FIRST-ORDER LOW-PASS FILTER

The impulse response of the corresponding time-continuous system is given with Equation 2.99. From this, we have

$$h[n] = \begin{cases} ae^{-anT} & n \geq 0 \\ 0 & n < 0 \end{cases} \quad (2.124)$$

For an input  $x[m]$ , the response becomes

$$\begin{aligned} y[m] &= x[m] * h[m] \\ &= a \sum_{n=0}^{\infty} x[m-n]b^n \quad (b = e^{-aT}) \end{aligned}$$

If a rectangular pulse  $x[n] = \begin{cases} 1 & 0 \leq n \leq N \\ 0 & \text{otherwise} \end{cases}$  is applied to the system, the output  $y[m]$  becomes the sampled version of the result shown in Figure 2.24.

**The Fourier Transform of Discrete-Time System** The sampled signal at the output of a linear time-invariant system is expressed as

$$\begin{aligned} y_s(t) &= \sum_{m=-\infty}^{\infty} y(mT)\delta(t-mT) \\ &= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(mT-nT)h(nT)\delta(t-mT) \quad (\text{from Eq. 2.123}) \end{aligned}$$

Taking the Fourier transform of the above equation, we have

$$Y_s(\omega) = H_s(\omega)X_s(\omega)$$

where,

$$\begin{aligned} H_s(\omega) &= \sum_{n=-\infty}^{\infty} h(nT)e^{-jn\omega T} \\ X_s(\omega) &= \sum_{n=-\infty}^{\infty} x(nT)e^{-jn\omega T}. \end{aligned}$$

Letting  $z = e^{j\omega T}$ , we have  $Y(z) = H(z)X(z)$ . Where  $H(z) = \sum_{n=-\infty}^{\infty} h[n]z^{-n}$ ,  $X(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n}$ , and  $Y(z) = \sum_{n=-\infty}^{\infty} y[n]z^{-n}$ , and the subscript  $s$  is omitted.

Since a sampled system has a gain of  $1/T$ , a constant  $T$  (sampling period) should be multiplied to the transfer function  $H_s(\omega)$  or  $H(z)$  to equalize the output signal level to that of the time-continuous system.

**Frequency Response of Discrete-Time System** We consider a complex sinusoidal signal  $x[m] = e^{jm\omega T}$  ( $-\infty \leq m \leq \infty$ ) as an input. Then from Equation 2.123, we have

$$\begin{aligned} y[m] &= \sum_{n=-\infty}^{\infty} e^{j(m-n)\omega T} h[n] \\ &= e^{jm\omega T} \sum_{n=-\infty}^{\infty} h[n]e^{-jn\omega T} \\ &= H(\omega)x[m] \end{aligned}$$

where we put  $H(\omega) = \sum_{n=-\infty}^{\infty} h[n]e^{-jn\omega T}$ .  $H(\omega)$  is called frequency transfer function or system function of a time-discrete system. The output  $y[m]$  is given merely multiplying  $H(\omega)$  to the input  $x[m]$ .  $H(\omega)$  and  $x[m] = e^{jm\omega T}$  are called in the convolution the *eigenvalue* and the *eigenfunction*, respectively.

If we put  $z = e^{j\omega T}$  for the system function, we have

$$H(z) = \sum_{n=-\infty}^{\infty} h[n]z^{-n}.$$

### EXAMPLE 2.14 DELAY LINE ( $H[N] = \Delta[N - N]$ )

From Equation 2.114, we have the transfer function as

$$H(\omega) = e^{-j\omega NT}, \quad H(z) = z^{-N}.$$

If we denote the input and output signals as  $x[n] \leftrightarrow X(z)$  and  $y[n] \leftrightarrow Y(z)$ , we get  $Y(z) = z^{-N}X(z) \leftrightarrow y[n] = x[n - N]$ .

### EXAMPLE 2.15 FIRST-ORDER LOW-PASS FILTER

Taking the discrete-time Fourier transform and  $z$ -transform of Equation 2.124, we have

$$H(\omega) = T \sum_{n=0}^{\infty} ae^{-naT} e^{-jn\omega T} = \frac{aT}{1 - e^{-aT} e^{-j\omega T}} \quad (a \geq 0) \quad (2.125)$$

$$H(z) = T \sum_{n=0}^{\infty} ae^{-naT} z^{-n} = T \sum_{n=0}^{\infty} ab^{-n} z^{-n} = \frac{aT}{1 - b^{-1} z^{-1}} \quad (a: \text{arbitrary}, \quad b = e^{aT}) \quad (2.126)$$

where we introduced the constant  $T$  (sampling period) to equalize the gain to that of the time-continuous system. The transfer function differs from that  $H(\omega) = 1/(1 + j\omega/a)$  for the time-continuous system. This is because of that the sampling theorem never holds, since  $H(\omega)$  spreads into a finite region. If the sampling frequency is sufficiently high and  $aT \ll 1$ ,  $|\omega T| \ll 1$ , the transfer function approaches to that of the time-continuous system as  $H(\omega) \approx 1/(1 + j\omega/a)$ .

**Stability of Discrete-Time System** A system is stable if for any bounded input signal  $x[n]$ , the output  $y[n]$  is bounded. The necessary and sufficient condition for the stability of the system becomes

$$\sum_{n=-\infty}^{\infty} |h[n]| < \infty \quad (2.127)$$

*Proof: Sufficient condition* We can show that

$$\begin{aligned} |y[n]| &= \left| \sum_{m=-\infty}^{\infty} x[n-m]h[m] \right| \\ &\leq \sum_{m=-\infty}^{\infty} |x[n-m]| |h[m]| \leq M \sum_{m=-\infty}^{\infty} |h[m]| < \infty \end{aligned}$$

where  $M = \max \{|x[n]|\}$ .

*Necessary condition* We assume that the condition with Equation 2.127 is unnecessary, that is, the system is stable even for  $\sum_{n=-\infty}^{\infty} |h[n]| = \infty$  (proof by contradiction).

We consider a bounded input signal  $x[n] = h^*[N-m] / |h[N-m]|$ , then the output  $y[n]$  becomes

$$\begin{aligned} y[n] &= \sum_{m=-\infty}^{\infty} h[m]x[n-m] \\ &= \sum_{m=-\infty}^{\infty} h[m] \frac{h^*[N-(n-m)]}{|h[N-(n-m)]|} \end{aligned}$$

For  $n = N$ , we have

$$y[N] = \sum_{m=-\infty}^{\infty} |h[m]| = \infty$$

Thus, the system is never stable. Therefore, the condition is necessary.

## EXAMPLE 2.16 FIRST-ORDER LOW-PASS FILTER

The impulse response is given with Equation 2.124. Then, we have

$$\sum_{n=-\infty}^{\infty} |h[n]| = a \sum_{n=0}^{\infty} e^{-naT} = \begin{cases} \frac{1}{1-e^{-aT}} & e^{-aT} < 1 \\ \infty & e^{-aT} > 1 \end{cases}$$

Thus, the system is stable if  $a > 0$ .

## EXAMPLE 2.17

$$h(nT) = \begin{cases} nT e^{-anT} & (a > 0) \quad n > 0 \\ 0 & n \leq 0 \end{cases} \quad (2.128)$$

From the mathematical formula,

$$\sum_{n=1}^{\infty} nx^n = x \frac{d}{dx} \sum_{n=1}^{\infty} x^n = \frac{x}{(1-x)^2} \quad (|x| < 1) \quad (2.129)$$

if  $x = e^{-aT} < 1$ , in other words,  $a > 0$ , we have  $\sum_{n=-\infty}^{\infty} |h[n]| = \frac{Te^{-aT}}{(1-e^{-aT})^2} < \infty$  and the system is stable. From the formula  $f_{k+1}(x) = xf'_k(x)$  for  $\sum_{n=1}^{\infty} n^k x^n = f_k(x)$  ( $|x| < 1$ ), the system with

$$h(nT) = \begin{cases} (nT)^k e^{-anT} & (a > 0) \quad n > 0 \\ 0 & n \leq 0 \end{cases}$$

can be shown to be also stable if  $a > 0$ .

Now we investigate the stability when the system function is given by a rational function as

$$H(z) = \frac{\sum_{n=0}^N a_n z^{-n}}{\sum_{m=0}^M b_m z^{-m}}$$

This is a case when the system is described with a difference equation as discussed in the following section. The polynomial in the denominator expresses the feedback part of the system and therefore it governs the stability of the system. Factorizing it, we express it as  $C \prod_{m=1}^M (1 - \beta_m z^{-1})$ . Where  $\beta_m$  (poles) are the roots of the denominator polynomial of  $z$ , that is, they satisfy  $1 - \beta_m z^{-1} = 0$ . The roots  $b_m$  ( $m = 1, 2, \dots, M$ ) of  $\sum_{m=0}^M b_m z^{-m} = 0$  are real or the complex conjugate roots, since the coefficients  $b_m$  (and  $a_n$ ) are real. Here, we assume that the poles are different from zeros (the roots of the decimator polynomial): if some of them are equal, we can reduce the polynomial. The polynomial  $\frac{1}{B(z)} = \frac{1}{C \prod_{m=1}^M (1 - \beta_m z^{-1})}$  can be decomposed into partial fractions as  $\sum_{m=1}^M \frac{\gamma_m}{1 - \beta_m z^{-1}}$ . The  $z$ -transform  $H_m(z) = \frac{\gamma_m}{1 - \beta_m z^{-1}}$  of each partial fraction is the same as the one with Equation 2.126 and its impulse response becomes

$$h_n(n) = \begin{cases} r_m e^{-a_m nT} & n \geq 0 \\ 0 & n < 0 \end{cases}$$

where  $\beta_m = e^{-a_m T}$ . From the example given with Equation 2.124, the system with this impulse response is stable if  $a_m > 0$  or  $|\beta_m| < 1$ . Thus, the system is stable if  $|\beta_m| < 1$  for all of the roots. We assumed simple roots so far. Even for multiple roots we have the same conclusion as follows: (the  $z$ -transform of an impulse response with Equation 2.128)



$$\begin{aligned}
 H(z) &= \sum_{n=1}^{\infty} nT e^{-anT} z^{-n} \\
 &= T \frac{e^{-aT} z^{-1}}{(1 - e^{-aT} z^{-1})^2} \quad (\text{from Eq. 2.129})
 \end{aligned} \tag{2.130}$$

$H(z)$  has a double root  $e^{-aT}$ . Since  $|e^{-aT}| < 1$  for  $a > 0$ , the stability of the system is the same as the one with a simple root. We can see that a system with its impulse response

$$h[n] = \begin{cases} (nT)^k e^{-anT} & n \geq 0 \\ 0 & n < 0 \end{cases}$$

is stable if  $a > 0$ . The  $z$ -transform of the impulse response has a  $k+1$  multiple root and therefore the stability of a system with a multiple root can be judged same as the one with a simple root.

### EXAMPLE 2.18

We find impulse response and investigate the stability of a system with transfer function as

$$H(z) = \frac{1}{1 - (a + a^*)z^{-1} + |a|^2 z^{-2}} \tag{2.131}$$

(i)  $a \neq a^*$  :  $H(z)$  can be modified as

$$H(z) = \frac{\frac{a^*}{a^* - a}}{1 - \frac{a^*}{a} z^{-1}} + \frac{\frac{a}{a - a^*}}{1 - a z^{-1}}$$

From Equations 2.120 and 2.121, we have

$$h[n] = \begin{cases} \frac{a^*}{a^* - a} (a^*)^n + \frac{a}{a - a^*} (a)^n & n \geq 0 \\ 0 & n < 0 \end{cases}$$

From this, the system is stable if  $|a| < 1$ .

(ii)  $a = a^*$ :  $H(z)$  becomes

$$H(z) = \frac{1}{(1 - az^{-1})^2}$$

From Equations 2.128 and 2.130, we get

$$g[n] = \begin{cases} na^n & n \geq 0 \\ 0 & n < 0 \end{cases} \leftrightarrow G(z) = \frac{az^{-1}}{(1-az^{-1})^2}$$

Since  $G(z) = az^{-1}H(z)$ , we have  $g[n] = a\delta[n-1] * h[n] = ah[n-1]$ . Then, we get

$$\sum_{n=-\infty}^{\infty} |ah[n-1]| = \sum_{n=-\infty}^{\infty} |g[n]| = \begin{cases} < \infty & |a| < 1 \\ \rightarrow \infty & |a| \geq 1 \end{cases}$$

Thus, the system is stable if  $|a| < 1$ .

### 2.4.5 Description with Difference Equation

For an input  $x[n]$  and the output  $y[n]$ , a discrete-time (linear time-invariant) system is described with a difference equation as

$$\sum_{m=0}^M b_m y[i-m] = \sum_{n=0}^N a_n x[i-n] \quad (2.132)$$

where  $i$  denotes the present time. A way to obtain a differential equation is to approximate differentiation in a differential equation with difference as

$$\left. \frac{df(t)}{dt} \right|_{t=mT} \approx \frac{f(mT) - f(mT-T)}{T} \quad (2.133)$$

The similar approximation is applied to the higher order differentiation.

The difference equation can be solved recursively with given initial values  $y[-m]$  ( $m=0,1,2,\dots,M$ ). For a linear system, the initial values should be zero since for a zero input the output becomes zero. The impulse response is given by solving the equation for an input  $x[n] = \delta[n]$ .

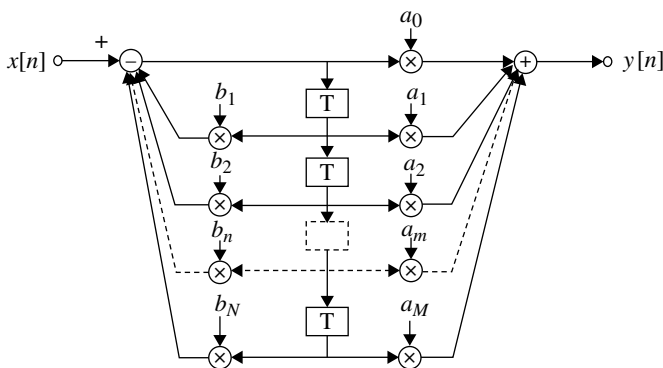
Letting the  $z$ -transform of  $y[n]$  and  $x[n]$  by  $Y(z)$  and  $X(z)$ , respectively, and taking the  $z$ -transform of Equation 2.132 (letting  $e^{j\Omega} = z$  in Eq. 2.114), we have

$$\sum_{m=0}^M b_m z^{-m} Y(z) = \sum_{n=0}^N a_n z^{-n} X(z)$$

Thus, we get

$$Y(z) = H(z)X(z)$$

where  $H(z) = \sum_{n=0}^N a_n z^{-n} / \sum_{m=0}^M b_m z^{-m}$  is the system function. The frequency transfer function  $H(\omega)$  is given by letting  $z = e^{j\omega T}$  (Section 2.4.4).



**FIGURE 2.35** Signal flow graph of a discrete-time system described with difference equation ( $M=N$ ,  $b_0 = 1$ ).

The signal flow graph is given by replacing differentiation with difference in the corresponding time-continuous system (Fig. 2.21). The delay circuits are commonly used by changing the order of connection of the feedback and feed forward parts (Fig. 2.35).

### EXAMPLE 2.19 FIRST-ORDER LOW-PASS FILTER

We consider a discrete-time system corresponding to a continuous-time system shown in Figure 2.22. From Equation 2.124, we have

$$\begin{aligned} y[m] &= \sum_{n=0}^{\infty} x[m-n]h[n] \\ &= a \sum_{n=0}^{\infty} x[m-n](e^{-aT})^n \end{aligned}$$

Then, we get the difference equation as

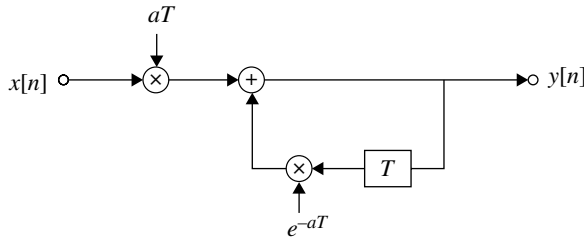
$$y[m] - e^{-aT} y[m-1] = ax[m] \quad (2.134)$$

Figure 2.36 shows the signal flow graph for this relation. We introduced a gain  $T$  to compensate the discrete-time system gain.

An easier method to get the signal flow graph is to take the  $z$ -transform of  $h(t)$  as

$$\begin{aligned} H(z) &= \sum_{n=0}^{\infty} h[n]z^{-n} = \sum_{n=0}^{\infty} a(e^{-aT} z^{-1})^n \\ &= \frac{a}{1 - e^{-aT} z^{-1}} \end{aligned}$$

Taking the  $z$ -transform of the difference equation with Equation 2.134 yields also the system function  $H(z)$  as



**FIGURE 2.36** Discrete-time first-order low-pass filter.

$$y(z) - e^{-aT} z^{-1} Y(z) = aX(z)$$

$$Y(z) = \frac{a}{1 - e^{-aT} z^{-1}} X(z) = H(z) X(z)$$

Figure 2.36 is readily given from the definition of  $H(z)$  and considering  $z^{-1} = e^{-j\omega T} \leftrightarrow \delta(t - T)$ .

Approximating differentiation with difference using Equation 2.133 in the differential equation, we have

$$y[m] - \frac{1}{1 + aT} y[m-1] = \frac{aT}{1 + aT} x[m]$$

For  $aT \ll 1$ , we get the approximation  $e^{-aT} \approx 1 - aT$ . Then, the above equation is further approximated as

$$y[m] - (1 - aT)y[m-1] = aTx[m]$$

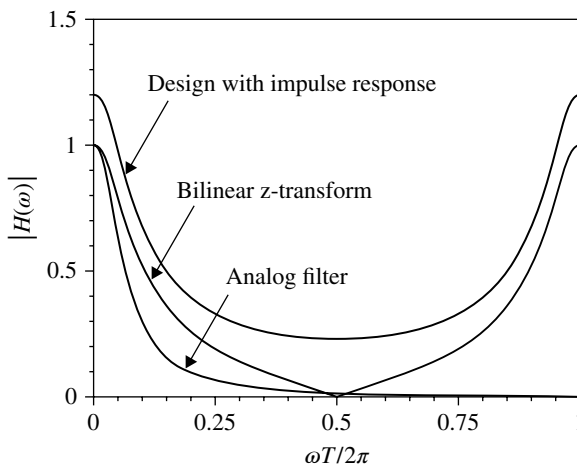
This equation differs from the previously obtained one (Eq. 2.134) only in the gain  $T$ .

## 2.4.6 Digital Filter

A filter realized with discrete-time signal processing is called digital filter. The elements of the filter are unit time delay elements, multiplexing circuits and adding (subtracting) circuits. An example of a digital filter is already shown in Figure 2.36. For a filter to be physically realizable, the impulse response should be real and the causality should hold. For a filter to be feasible, the stability as a linear system is required.

Characteristics required for a filter are described by its impulse response in time domain or the transfer function in frequency domain. A method to get a digital filter is to sample the impulse response of the corresponding analog (continuous-time) filter. Then, the transfer function is given similarly to Equation 2.108 as

$$H_s(\omega) = \frac{1}{T} \sum_{n=-\infty}^{\infty} H(\omega - n\omega_s) \quad (\omega_s = 2\pi/T, \quad T : \text{sampling period})$$



**FIGURE 2.37** Amplitude transfer function of first-order low-pass filters ( $aT = \pi/8$ ).

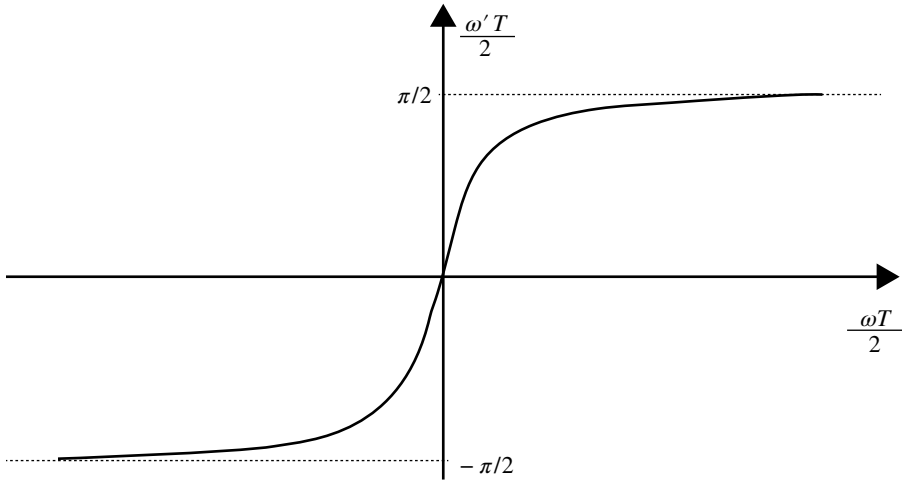
where  $H(\omega)$  is the transfer function of the analog filter. To achieve the analog filter characteristics precisely, the sampling theorem should be satisfied. To this end, the bandwidths of the input signal and the filter should be limited in a finite value. Otherwise error occurs due to the aliasing effect. In other words, the impulse response is not restored precisely at times except for the sampling instants. As discussed in Section 2.3.4 for example, the characteristics of an analog first-order low-pass filter cannot be restored precisely, since the frequency transfer function spans into an infinite region, although it decays as the frequency increases (Fig. 2.37). The error decreases if the sampling frequency becomes sufficiently high. Due to the same reason, no ideal high-pass filter can be realized. Therefore, suppression of aliasing error becomes an important issue when analog filter characteristics are realized with the digital filter. To this end, a method with frequency transformation is known as

$$\frac{\omega T}{2} = \tan \frac{\omega' T}{2} \quad (2.135)$$

With this transformation, the entire frequency region  $|\omega| \leq \infty$  is mapped into a finite region of  $|\omega'| < \pi/T$  and the aliasing error never occurs (Fig. 2.38). The frequency characteristics, however, deviate from the original ones. To make the characteristics closer to the original one, the transfer function  $H(\omega)$  is usually adjusted beforehand, for example, by controlling the cut-off frequencies to be the same. When  $|\omega' T| \ll 1$ , from Equation 2.135, we have  $\omega \approx \omega'$  and the characteristics are kept almost the same.

Letting  $z = e^{-j\omega' T}$ , Equation 2.135 is rewritten as

$$j\omega = \frac{2}{T} \frac{1 - z^{-1}}{1 + z^{-1}} \quad (2.136)$$



**FIGURE 2.38** Frequency transformation with bilinear  $z$ -transform.

The transformation is called bilinear  $z$ -transform from the expression.

Transfer function  $G(z) = H(j\omega) \Big|_{j\omega = \frac{2}{T} \frac{1-z^{-1}}{1+z^{-1}}}$  of the digital filter is obtained through inserting the above equation with transfer function  $H(j\omega)$  of an analog filter.

### EXAMPLE 2.20 FIRST-ORDER LOW-PASS FILTER

We will obtain digital filters from an analog first-order low-pass filter with its transfer function given as

$$H(\omega) = \frac{1}{1 + j\omega / a}$$

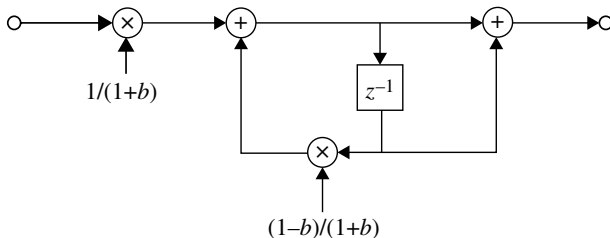
(i) *Design by impulse response:* From Equation 2.125, we have

$$H(z) = \frac{aT}{1 - e^{-aT} z^{-1}}$$

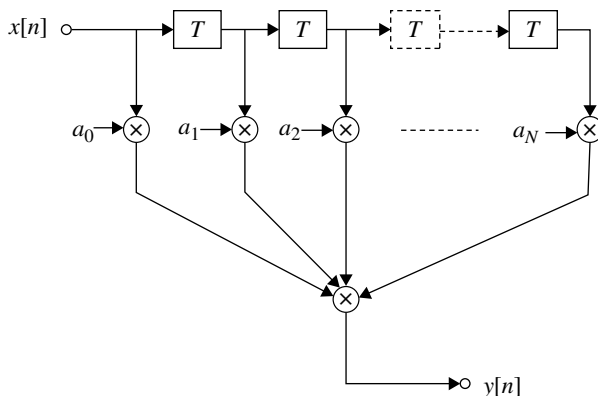
Even at the maximum frequency  $\omega = \pi/T$  ( $z^{-1} = -1$ ) which is able to be expressed in a discrete-time system, we have  $|H(z)| \neq 0$  (Fig. 2.37) due to the aliasing error. The circuit block diagram is shown in Figure 2.36.

(ii) *Design by bilinear  $z$ -transform:* We get

$$H(z) = \frac{1}{1 + \frac{2}{aT} \frac{1-z^{-1}}{1+z^{-1}}} = \frac{1+z^{-1}}{(1+b) \left[ 1 + \frac{1-b}{1+b} z^{-1} \right]} \quad \left( b = \frac{2}{aT} \right)$$



**FIGURE 2.39** First-order low-pass filter designed with the bilinear  $z$ -transform.



**FIGURE 2.40** Transversal filter.

At the frequency  $\omega = \pi/T$ , we have  $|H(z)| = 0$  due to no aliasing error (Fig. 2.37). Figure 2.39 shows the circuit block diagram.

**2.4.6.1 Finite Impulse Response Filter** This filter shows an impulse response that spans a finite time region. It should never include a feedback part (Fig. 2.40). The filter is also called transversal filter. The impulse response and the transfer function become

$$h(n) = \begin{cases} a_n & n = 0, 1, \dots, N \\ 0 & \text{otherwise} \end{cases}$$

and

$$H(z) = \sum_{n=0}^N a_n z^{-n}$$

This circuit is always stable due to no feedback part.

### EXAMPLE 2.21 MOVING AVERAGE CIRCUIT

Letting  $a_0 = a_1 = \dots = a_N = 1$ , we have  $y[n] = \sum_{m=0}^N x[n-m]$  which means the moving average of  $x[n]$ .

### EXAMPLE 2.22 LINEAR PHASE CIRCUIT

For a circuit with frequency transfer function  $H(\omega) = H(z)_{z=e^{j\omega T}}$ , it is called a linear phase circuit if the phase  $\angle H(\omega)$  is a linear function of  $\omega$ . The condition for the linear phase circuit is that its impulse response becomes symmetrical with respect to a center time, that is,

$$a_n = a_{N-n} \begin{cases} N : \text{even}, & n = 0, 1, \dots, N/2 - 1 \\ N : \text{odd}, & n = 0, 1, \dots, (N-1)/2 \end{cases}$$

The condition is confirmed as follows: Consider a case for  $N$  is even ( $N=2M$ ). We have

$$\begin{aligned} H(\omega) &= \sum_{n=0}^N a_n e^{-jn\omega T} \\ &= \sum_{n=0}^{M-1} \left[ a_n e^{-jn\omega T} + a_{N-n} e^{-j(N-n)\omega T} \right] \quad (a_{N-n} = a_n) \\ &= \sum_{n=0}^{M-1} a_n e^{jN\omega T/2} \left[ e^{j(N/2-n)\omega T} + e^{-j(N/2-n)\omega T} \right] \\ &= e^{-jN\omega T/2} \sum_{n=0}^{M-1} a_n \cos(N/2 - n)\omega T \end{aligned}$$

Thus,  $\angle H(\omega) = -\omega NT/2$  is shown. The similar argument holds for an odd  $N$ .

**2.4.6.2 Infinite Impulse Response Filter** A filter with its impulse response that spans infinitely is called IIR filter. The filter includes a feedback part and possibly becomes unstable. An example of the IIR digital filter impulse response is shown with Equation 2.124.

In designing a digital filter to approximate a given characteristic, the number of filter elements can be decreased by including an IIR filter part. For example, many elements are required for an FIR filter to realize the impulse response given with Equation 2.124. In contrast to this, the IIR filter is implemented simply as shown in Figure 2.39.

### 2.4.7 Downsampling, Upsampling, and Subsampling

A sampling frequency is chosen usually higher than the minimum frequency required by the sampling theorem. It is not preferable for us to perform signal processing such as filtering at that sampling frequency in view of the operation complexity. Decreasing the sampling frequency is called downsampling. It is easy to decrease the sampling frequency by a factor of an integer number: we merely pick the sampled values at every  $N$  samples. An upsampling with a factor of an integer number  $M$  is carried out by inserting  $M-1$  zeros between the successive two samples. Upsampling never changes the signal spectrum as explained as follows: we start our discussion in



frequency domain for simplicity. Consider a signal  $f(t)$  and its Fourier transform  $F(\omega)$ . We make a spectrum by placing  $F(\omega)$  regularly  $M$ -times at a spacing of  $\omega_0$ , that is,

$$F_M(\omega) = \sum_{m=0}^{M-1} F(\omega - m\omega_0)$$

Taking the inverse transform of the above equation, we have

$$f_M(t) = f(t) \sum_{m=0}^{M-1} e^{jm\omega_0 t}$$

Here we consider  $f_M(t)$  at times  $t = i\Delta t$ , where  $\Delta t = 2\pi / (M\omega_0)$ . Then, we get

$$f_M(i\Delta t) = f(i\Delta t) \sum_{m=0}^{M-1} e^{j\frac{mi}{M}2\pi}$$

Letting  $i = lM + i'$  ( $i' = 0, 1, 2, \dots, M-1$ ,  $l = 0, 1, 2, \dots$ ), we have

$$f_M(i\Delta t) = f(i\Delta t) \sum_{m=0}^{M-1} e^{j\frac{mi'}{M}2\pi}$$

From this we get  $f_M(i\Delta t) = Mf(i\Delta t)$  for  $i' = 0$  and  $f_M(i\Delta t) = 0$  for  $i' \neq 0$ . Where we used

$$\sum_{m=0}^{N-1} e^{-j2\pi nm/N} = \begin{cases} N & n = 0 \\ 0 & n \neq 0 \end{cases}$$

From the aforementioned argument, it is shown that the insertion of  $M - 1$  zeros between the neighboring samples in the time domain repeats the original spectrum  $M$  times in the frequency domain. The highest frequency to be treated in the sampling system increases by  $M$ -times owing to the upsampling. For example, we can use a digital low-pass filter to suppress the higher harmonics of the spectrum. Then, the sampled signals of the original analog signal with the  $M$ -times higher sampling frequency is obtained at the output of the digital low-pass filter. Then, we can relax the frequency transfer characteristics of an analog interpolation filter to convert the digital signal to the analog signal. This technique with the upsampling and the digital low-pass filter is actually used for CD (compact disc) players for an example.

Conversely, if we put  $M - 1$  zero spectra in frequency domain,  $M$  signals repeat in time domain as follows: We investigate whether spectrum of the signal constituting of a signal  $f(t)$  is repeated with a spacing of  $T$  in time domain. The signal is described as

$$f_M(t) = \sum_{m=0}^M f(t - mT)$$

Taking the Fourier transform of the signal, we have

$$f_M(t) = \sum_{m=0}^M f(t - mT)$$

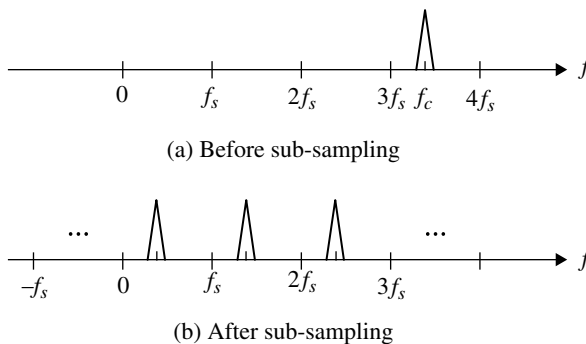
We consider the spectrum at frequencies  $\omega = i\Delta\omega$ , where  $\Delta\omega = \omega_0 / M$  and  $\omega_0 = 2\pi / (MT)$  are understood. Letting  $i = lM + i'$  ( $i' = 0, 1, 2, \dots, M-1, l = 0, 1, 2, \dots$ ), we can show that

$$\begin{aligned} F_M(i\Delta\omega) &= MF(i\Delta\omega) \quad (i' = 0) \\ F_M(i\Delta\omega) &= 0 \quad (i' \neq 0) \end{aligned}$$

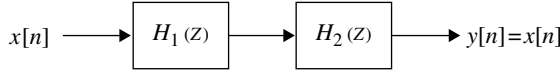
From the reciprocity of the Fourier transform, it is shown that inserting  $M-1$  zero spectra in frequency domain for a given signal at the spacing of  $\Delta\omega$  yields a signal repeating the given signal  $M$ -times in time domain. This fact is used for a single-carrier multiplexing system, where different  $M$ -signals are orthogonally multiplexed by inserting  $M$  spectra components periodically (interleaving) in frequency domain (Section 6.6) keeping the single-carrier characteristics for each multiplexed signal: The frequency components of each signal is distributed periodically and therefore its bandwidth is spread by  $M$ -times.

To change the sampling frequency from  $N$  to  $M$ , we take a process as follows: First, we apply upsampling (zero insertion) with a sampling frequency of the least common multiple ( $L$ ) of  $M$  and  $N$ . Then, interpolation is done with a digital filter. Downsampling of the interpolated signal with a factor of  $L/N$  yields the desired signal.

**Subsampling** Next, we consider a narrow band bandpass signal, such as an (IF) signal, which has a narrow (double-sided) spectrum bandwidth  $2B$  around a center frequency  $f_c$  ( $>2B$ ) (Fig. 2.41a). In this case, we need not take a high sampling frequency (at least higher than  $2f_c$ ) to meet the sampling theorem but can use a sampling frequency of  $f_s$  higher than  $2B$ . This is called subsampling. The sampled signal spectrum becomes as shown in Figure 2.41b. This is the same as that for a low-pass signal. However, in an actual system, a little difference occurs due to



**FIGURE 2.41** Spectrum in subsampling.

**FIGURE 2.42** Inverse circuit.

fluctuation of the sampling frequency (and hence the phase) called jitter. The jitter becomes high proportionally to the order  $m$  of the higher harmonics  $mf_s$ . In the case of Figure 2.41b, the baseband signal component is produced with the third-order higher harmonics ( $3f_s$ ). The jitter becomes significant if the sampling frequency is taken to be too low, since a very high harmonic produces the baseband signal component.

### 2.4.8 Inverse Circuit

For a given circuit with transfer function  $H_1(z)$ , when the other circuit with  $H_2(z)$  is serially connected, if the overall transfer function becomes  $H(z) = H_1(z)H_2(z) = 1$  (or  $z^{-m}$ ), the circuit is called inverse circuit (Fig. 2.42). The impulse response becomes  $h[n] = h_1[n] * h_2[n] = \delta[n]$  (or  $\delta[n - m]$ ).

#### EXAMPLE 2.23 $H_1(z) = a + bz^{-1}$

The transfer function of the inverse circuit becomes

$$H^{-1}(z) = \frac{1}{a + bz^{-1}}$$

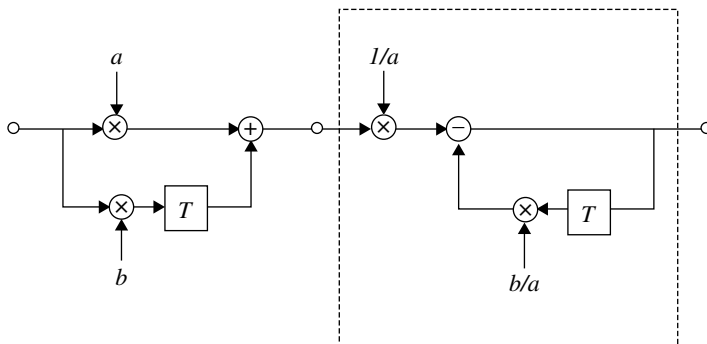
The circuit is stable for  $|b/a| < 1$ . Expressing as  $H^{-1}(z) = \frac{1}{a} \frac{1}{1 + z^{-1}b/a} = \sum_{n=0}^{\infty} \frac{1}{a} \left(-\frac{b}{a}\right)^n z^{-n} \left(\frac{b}{a} < 1\right)$ , we have the impulse response as

$$h[n] = \begin{cases} \frac{1}{a} \left(-\frac{b}{a}\right)^n & n \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

This circuit is implemented as shown within the broken line area in Figure 2.43. The circuit is an IIR filter, since it includes a feedback part. We can confirm directly in the figure that the output becomes the same as the input ( $y[n] = x[n]$ ).

### 2.4.9 Window Function

We consider to approximate an impulse response of an IIR filter with an FIR filter. The realized frequency response spreads into an infinite region, even when the original frequency response is limited in a finite region, since the realized



**FIGURE 2.43** An example of inverse circuit connection.

(approximated) IIR filter has a FIR. To suppress the effect, a time (window) function with a finite time length is multiplied with the impulse response, that is,

$$h_w[n] = h[n]w[n] \quad (|n| = 0, 1, \dots, (N-1)/2, \quad N : \text{odd})$$

where  $w[n] = 0$  ( $|n| > (N-1)/2$ ).

### EXAMPLE 2.24 RECTANGULAR WINDOW

The window function is given as

$$w[n] = \begin{cases} 1 & |n| \leq (N-1)/2 \\ 0 & \text{otherwise} \end{cases}$$

This corresponds to merely truncate the impulse response  $h[n]$  in  $|n| \leq (N-1)/2$ .

### EXAMPLE 2.25 THE GENERALIZED HAMMING WINDOW

This window function becomes

$$w[n] = \begin{cases} \alpha + (1-\alpha) \cos 2\pi n / (N-1) & |n| \leq (N-1)/2 \\ 0 & \text{otherwise} \end{cases}$$

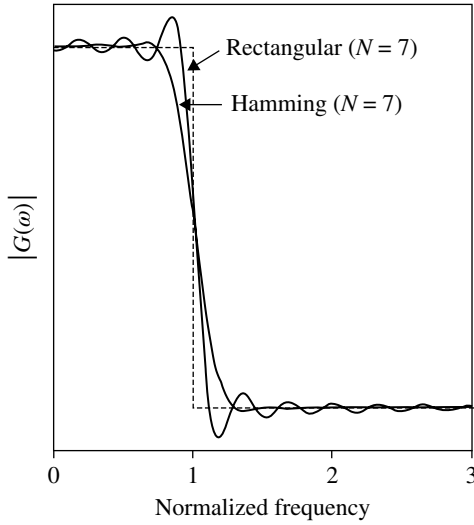
where  $0 \leq \alpha \leq 1$ . We call it the Hanning window when  $\alpha = 0.54$  and the Hamming window when  $\alpha = 0.5$ .

Amplitude transfer functions of an ideal low-pass filter with those window functions are shown in Figure 2.44.

### 2.4.10 Discrete Fourier Transform

We will get the spectrum of a signal  $f(t)$  with its sampled value. Using the sampling interval of  $T_s$ , the sampled signal is given as

$$f_s(t) = f(t) \cdot \sum_{m=-\infty}^{\infty} \delta(t - mT_s) \quad (2.137)$$



**FIGURE 2.44** Effects of window functions.

$$= \sum_{m=-\infty}^{\infty} f(mT_s) \delta(t - mT) \quad (\text{from Eq. 2.9}) \quad (2.138)$$

Taking the Fourier transform of Equation 2.138, we have

$$F_s(\omega) = \sum_{m=-\infty}^{\infty} f(mT_s) e^{-j\omega mT_s}$$

This is a periodic function of  $\omega$  with a period of  $\omega_s = (2\pi/T_s)$ . To focus on the periodicity of the function  $e^{-j\omega mT_s}$ , we introduce the variables  $T_p = NT_s$  and  $m = kN + n$ , and then we have

$$F_s(\omega) = \sum_{n=0}^{N-1} \sum_{k=-\infty}^{\infty} f(kT_p + nT_s) e^{-j\omega(kN+n)T_s} \quad (2.139)$$

If  $f(t)$  is defined in a finite time region,  $F_s(\omega)$  can be obtained with a finite number calculation.

Now we consider  $F_s(\omega)$  at frequencies  $\omega = 2\pi m/T_p \equiv m\Omega$  ( $\Omega = 2\pi/T_p$ ) ( $m = 0, \pm 1, \pm 2, \dots$ ). Then, we get

$$\begin{aligned} F_s(m\Omega) &= \sum_{n=0}^{N-1} \sum_{k=-\infty}^{\infty} f(kT_p + nT_s) e^{-j2\pi mn/N} \\ &= \sum_{n=0}^{N-1} e^{-j2\pi mn/N} f_p(nT_s) \end{aligned} \quad (2.140)$$

where  $f_p(t) = \sum_{k=-\infty}^{\infty} f(t + kT_p)$  ( $0 < t < T_p$ ).

Solving Equation 2.140 for  $m = 0, 1, 2, \dots, N-1$ , we have

$$f_p(nT_s) = \frac{1}{N} \sum_{m=0}^{N-1} F_s(m\Omega) e^{j2\pi mn/N} \quad (n = 0, 1, 2, \dots, N-1)$$

The following method is convenient to get this result. Multiplying the both sides of Equation 2.140 with  $e^{j2\pi mk/N}$  and taking the summation for  $m$ , we get

$$\begin{aligned} \sum_{m=0}^{N-1} e^{j2\pi mk/N} F_s(m\Omega) &= \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} e^{j2\pi mk/N} e^{-j2\pi mn/N} f_p(nT_s) \\ &= \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} e^{-j2\pi m(n-k)/N} f_p(nT_s) \\ &= N f_p(kT_s) \quad (k = 0, 1, \dots, N-1) \end{aligned}$$

where we used the following relation:

$$\sum_{m=0}^{N-1} e^{-j2\pi mn/N} = \begin{cases} N & n = 0 \\ 0 & m \neq 0 \end{cases}$$

Since  $F_s(m\Omega)$  is a periodic function with a period of  $N\Omega$ , we have the following two relations:

$$F_s(m\Omega) = \sum_{n=0}^{N-1} f_p(nT_s) e^{-j2\pi mn/N} \quad (m = 0, 1, \dots, N-1) \quad (\text{DFT}) \quad (2.141)$$

$$f_p(nT_s) = \frac{1}{N} \sum_{m=0}^{N-1} F_s(m\Omega) e^{j2\pi mn/N} \quad (n = 0, 1, \dots, N-1) \quad (\text{IDFT}) \quad (2.142)$$

These relations are called the discrete Fourier transform (DFT) of  $f_p(nT_s)$  and the inverse discrete Fourier transform (IDFT) of  $F_s(m\Omega)$ , respectively.

In the DFT, the maximum frequency becomes  $(N-1)\Omega = (1-1/N)\omega_s$  and the sampling theorem seems unsatisfied. However, since only positive frequencies are considered using complex signal expression, the sampling theorem is satisfied (Section 2.4.1).

The function  $f_p(t)$  is arbitrary. Therefore, Equation 2.141 can be considered to show the Fourier transform  $F_s(\omega)$  for  $\omega = m\Omega$  ( $m = 0, 1, \dots, N-1$ ) of a signal  $f_s(t) = \sum_{n=0}^{N-1} f(nT_s) \delta(t - nT_s)$  that is given by sampling an arbitrary function  $f(t)$  for  $N$ -times at a period of  $T_s$ , where  $\Omega = \omega_s/N$ ,  $\omega_s = 2\pi/T_s$ . Usually Equations 2.141 and 2.142 are understood in this sense. Sometimes, we want to get frequency components at  $\omega = m\Omega$  ( $m = 0, 1, \dots, N-1$ ) of a long-time continuing signal. In this case, it is not efficient to get average of each frequency components given by proceeding to a number of the DFTs of a block of  $N$  sampled signals successively. In contrast to this, one-time DFT is sufficient if we take the DFT of  $N$  signals each of which is given by summation of the signals sampled in every  $N$ -sample time as Equation 2.141 originally means.

Now we consider what  $f_p(nT_s)$  and  $F_s(m\Omega)$  mean. Taking the Fourier transform of Equation 2.137 from Equations 2.40 and 2.18, we have an equation, which is a different expression derived from Equation 2.138, and is as follows:

$$F_s(\omega) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} F(\omega - n\omega_s)$$

If  $F(\omega) = 0$  for  $|\omega| > \omega_s / 2$ , that is, the sampling theorem holds, we have

$$F_s(m\Omega) = \frac{1}{T_s} F(m\Omega) \quad (|m\Omega| \leq \omega_s)$$

and  $F(m\Omega)$  is obtained exactly with Equation 2.141. Otherwise the harmonic spectra components overlap with each other and the aliasing error occurs.

When  $f(t)$  is constrained within a time  $T_p$ , that is,  $f(t) = 0$  ( $t < 0$  or  $t > T_p$ ), then we have  $f_p(t) = f(t)$  ( $0 \leq t \leq T_p$ ). Thus, Equation 2.141 becomes the Fourier transform of  $f(t)$ . In this case, the aliasing error cannot be avoided since the spectrum  $F(\omega)$  spreads into an infinite frequency range due to the fact that the signal  $f(t)$  is time limited (Section 2.1.2.6). To suppress the aliasing error, a higher sampling frequency  $\omega_s$  should be taken. The spectrum of a time-limited signal becomes continuous (Section 2.4.1). The continuous spectrum is obtained through interpolation of discrete components  $F(m\Omega)$  (the frequency sampling theorem, Section 2.4.1). The maximum frequency to be able to express becomes one-half of the sampling frequency with DFT of a real signal.

Next, we consider a case when  $f(t)$  is a periodic signal with a period of  $T_p$ . Then, the sampling theorem might hold, since the signal is not time-limited. The signal  $f(t)$  is given as

$$f(t) = \sum_{m=-\infty}^{\infty} f_T(t - mT_p) \quad (2.143)$$

where

$$f_T(t) = \begin{cases} f_0(t) & 0 \leq t \leq T_p \\ 0 & \text{otherwise} \end{cases}$$

Substituting Equation 2.143 into 2.139, we have

$$\begin{aligned} F_s(\omega) &= \sum_{n=0}^{N-1} \sum_{k=-\infty}^{\infty} f(kT_p + nT_s) e^{-j\omega(kN+n)T_s} \\ &= \sum_{n=0}^{N-1} f_0(nT_s) \sum_{k=-\infty}^{\infty} e^{-j\omega(kN+n)T_s} \\ &= \sum_{n=0}^{N-1} f_0(nT_s) e^{-j\omega nT_s} \sum_{k=-\infty}^{\infty} e^{-j\omega kT_p} \quad (T_p = NT_s) \end{aligned}$$

Referring to the derivation process of Equation 2.19, we can show that

$$\sum_{k=-\infty}^{\infty} e^{-j\omega k T_p} = \Omega \sum_{m=-\infty}^{\infty} \delta(\omega - m\Omega) \quad \left( \Omega = \frac{2\pi}{T_p} \right)$$

From this we have

$$F_s(\omega) = \Omega \sum_{m=-\infty}^{\infty} F_0(m\Omega) \delta(\omega - m\Omega) \quad (2.144)$$

$F_s(\omega)$  has a line spectrum, where  $F_0(\omega) = \sum_{n=0}^{N-1} f_0(nT_s) e^{-j\omega n T_s}$ . The magnitude of the line spectrum at  $\omega = m\Omega$  is given by integrating Equation 2.144 within a small region around  $m\Omega$  as follows:

$$\Omega F_0(m\Omega) = \Omega \sum_{n=0}^{N-1} f_0(nT_s) e^{-j2\pi mn/N}$$

The other way to get  $F_s(\omega)$  with Equation 2.144 is to take the Fourier transform of the periodic sampled signal.

If we use DFT and IDFT, the upsampling becomes easy: after taking DFT, we add frequency components with a null value at higher frequency part and take IDFT. The increase in the clock frequency corresponds to the increase in frequency components.

## 2.4.11 The Fast Fourier Transform

This is a method called fast Fourier transform (FFT) to calculate DFT fast. It becomes powerful, for example, when getting a signal spectrum or response of a linear system. We rewrite the DFT and IDFT given with Equations 2.141 and 2.142 as follows:

$$F[m] = \sum_{n=0}^{N-1} f[n] W_N^{mn} \quad (m = 0, 1, 2, \dots, N-1) \quad (\text{DFT}) \quad (2.145)$$

$$f[n] = \frac{1}{N} \sum_{m=0}^{N-1} F[m] W_N^{-mn} \quad (n = 0, 1, 2, \dots, N-1) \quad (\text{IDFT}) \quad (2.146)$$

where  $W_N = e^{-j2\pi/N}$ .

Here, we consider to calculate  $F[m]$  ( $m = 0, 1, 2, \dots, N-1$ ) (DFT). The calculation of  $f[n]$  ( $n = 0, 1, 2, \dots, N-1$ ) becomes similar to this. To calculate a  $F[m]$ ,  $N$ -times multiplication and  $N-1$  times addition are necessary. Calculation of  $F[m]$  for  $m = 0, 1, 2, \dots, N-1$  requires  $N^2$ -times multiplications. The FFT reduces the number of the multiplications to  $M \log_2 N$ . We can confirm the usefulness of the FFT by taking  $N=1024$ , for example, and having  $N^2 \approx 10^6$  and  $N \log_2 N \approx 10^4$ : the number of multiplications is reduced to be 1/100 of that for the DFT.

We assume  $N=2^M$  ( $M$ : integer). In Equation 2.145, we divide  $f[n]$  into two groups with even or odd  $n$  (decimation in time) as



$$\begin{aligned}
 F[m] &= \sum_{n=0}^{N/2-1} f[2n]W_N^{2nm} + \sum_{n=0}^{N/2-1} f[2n+1]W_N^{(2n+1)m} \\
 &= E[m] + W_N^m Q[m] \quad (m = 0, 1, 2, \dots, N-1)
 \end{aligned} \tag{2.147}$$

where

$$E[m] = \sum_{n=0}^{N/2-1} f[2n]W_N^{2nm}, \quad Q[m] = \sum_{n=0}^{N/2-1} f[2n+1]W_N^{2nm}$$

If  $E[m]$  and  $Q[m]$  are given, the number of multiplication and summation in Equation 2.147 becomes  $N$  ( $m = 0, 1, 2, \dots, N-1$ ). Using the relation  $W_N^{2nm} = W_{N/2}^{nm}$  and  $W_{N/2}^{n(N/2+m)} = W_{N/2}^{nm}$ ,  $E[m]$  and  $Q[m]$  are expressed with  $N/2$ -point DFT as follows:

$$\begin{aligned}
 E[m] &= E[N/2 + m] = \sum_{n=0}^{N/2-1} f[2n]W_{N/2}^{nm} \quad (m = 0, 1, \dots, N/2-1) \\
 Q[m] &= Q[N/2 + m] = \sum_{n=0}^{N/2-1} f[2n+1]W_{N/2}^{nm} \quad (m = 0, 1, \dots, N/2-1)
 \end{aligned}$$

Similarly  $E[m]$  and  $Q[m]$  are further modified as

$$\begin{aligned}
 E[m] &= \sum_{n=0}^{N/4-1} f[4n]W_{N/2}^{2nm} + W_{N/2}^m \sum_{n=0}^{N/4-1} f[4n+2]W_{N/2}^{2nm} \quad (m = 0, 1, 2, \dots, N/2-1) \\
 Q[m] &= \sum_{n=0}^{N/4-1} f[4n+1]W_{N/2}^{2nm} + W_{N/2}^m \sum_{n=0}^{N/4-1} f[4n+3]W_{N/2}^{2nm} \quad (m = 0, 1, \dots, N/2-1)
 \end{aligned}$$

The above  $E[m]$  or  $Q[m]$  are given by two times  $N/4$ -point DFT, a multiplication with  $W_{N/2}^m$  and a summation for a given  $m$ . When the  $N/4$ -point DFTs are given, calculation of  $E[m]$  or  $Q[m]$  for  $m = 0, 1, 2, \dots, N/2-1$  requires  $N/2$  multiplications and summations, and therefore the total number of multiplication and summation becomes  $N$ . Similar to the above argument,  $N/4$  operations (multiplication and summation) are done for 4 DFT groups and therefore  $N$  operations are performed in the next stage. These manipulations continue to the  $M$ th stage. The total overall number of operations becomes  $MN = N \log_2 N$ .

If we use the relations  $W_N^{2(m+N/2)} = W_N^{2m}$  and  $W_N^{m+N/2} = -W_N^m$ , Equation 2.147 can be rewritten as

$$\begin{aligned}
 F[m] &= E[m] + W_N^m Q[m] \quad (m = 0, 1, \dots, N/2-1) \\
 F[m + N/2] &= E[m] - W_N^m Q[m] \quad (m = 0, 1, \dots, N/2-1)
 \end{aligned}$$

From the above expression, the number of multiplications becomes  $(N/2) \log_2 N$  if the minus operation is not counted as multiplication. Figure 2.45 shows the FFT algorithm for  $N=8$  for an example. Eight times multiplications and summations are done at each stage. Each node diverges into two paths and for which the weighting factor becomes  $W_8^m$  or  $W_8^{m+4} = -W_8^m$ .

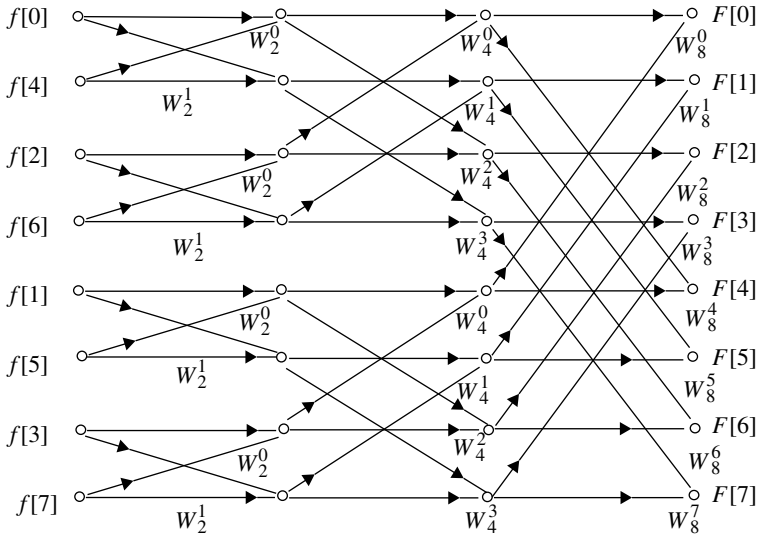


FIGURE 2.45 Example of FFT algorithm ( $N=8$ ).

From the similarity between DFT (Eq. 2.145) and IDFT (Eq. 2.146), the same method can be applied to IDFT. The principle of the FFT method is owing to properties of the weighting factor  $W_N^{mn}$ . The weighting factor  $W_N^{mn}$  is symmetrical with respect to variables  $m$  and  $n$ . Considering this fact, the other method where the decimation is made in frequency domain can be applied.

## 2.5 OPTIMIZATION AND ADAPTIVE SIGNAL PROCESSING

We discuss briefly the optimum solution for signal system and adaptive signal processing techniques.

### 2.5.1 Solution of Optimization Problem

For simplicity, we assume two variables  $x_1, x_2$  and find the optimum values  $\tilde{x}_1, \tilde{x}_2$  to maximize a function  $f(x_1, x_2)$ , which is differentiable and convex.

(1) *Without constraint*

We take the first-order approximation of  $f(x_1, x_2)$  around  $x_1, x_2$  as

$$f(x_1 + \Delta x_1, x_2 + \Delta x_2) \cong f(x_1, x_2) + \frac{\partial f}{\partial x_1} \cdot \Delta x_1 + \frac{\partial f}{\partial x_2} \cdot \Delta x_2 \quad (|\Delta x_1|, |\Delta x_2| \ll 1)$$

When  $x_1 \neq \tilde{x}_1, x_2 \neq \tilde{x}_2$ , if we take  $\Delta x_1$  and  $\Delta x_2$  such that  $\frac{\partial f}{\partial x_1} \Delta x_1 > 0$  and  $\frac{\partial f}{\partial x_2} \Delta x_2 > 0$ ,

we approach closer to the maximum value, since  $f(x_1 + \Delta x_1, x_2 + \Delta x_2) > f(x_1, x_2)$ . The optimum values  $\tilde{x}_1, \tilde{x}_2$  are found for  $x_1, x_2$  to satisfy the following equations:

$$\frac{\partial f}{\partial x_1} = 0, \quad \frac{\partial f}{\partial x_2} = 0.$$

**EXAMPLE 2.26**

Target function is expressed with square of absolute value of a complex number as

$$\begin{aligned} J &= |f(z)|^2 \\ &= |f(x, y)|^2 \quad z = x + jy \quad (x, y : \text{real}) \end{aligned}$$

The extreme values become the solution to the following equations

$$\frac{\partial}{\partial x} J = 0, \quad \frac{\partial}{\partial y} J = 0 \quad (2.148)$$

The expression becomes simple by introducing the operators as

$$\frac{\partial}{\partial z} = \frac{1}{2} \left( \frac{\partial}{\partial x} - j \frac{\partial}{\partial y} \right), \quad \frac{\partial}{\partial z^*} = \frac{1}{2} \left( \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} \right)$$

Then, the partial differentiations are expressed as

$$\frac{\partial}{\partial x} J = 2 \operatorname{Re} \left\{ \frac{\partial}{\partial z^*} J \right\}, \quad \frac{\partial}{\partial y} J = 2 \operatorname{Im} \left\{ \frac{\partial}{\partial z^*} J \right\}$$

Thus, Equation 2.148 becomes simple as

$$\frac{\partial}{\partial z^*} J = 0$$

Since the target function is expressed as  $J(z, z^*)$ , when we carry out  $\frac{\partial}{\partial z^*} J(z, z^*)$ , the partial differentiation often becomes simple by using the following relations.

$$\frac{\partial}{\partial z^*} z = 0, \quad \frac{\partial}{\partial z^*} z^* = 1$$

**EXAMPLE 2.27 LEAST MEAN SQUARE METHOD**

We consider a system when a complex transmit signal  $s$  is received with a channel gain of  $G$  and noise  $n$  is added. The problem is to find the optimum gain to minimize the mean squared error. The received signal is expressed as

$$r = Gs + n$$

The error is defined as

$$\varepsilon = s - \hat{s}$$

where

$$\hat{s} = wr = w(Gs + n)$$

The mean squared error  $J$  becomes

$$\begin{aligned} J &= \langle |\varepsilon|^2 \rangle = \langle |s - w(Gs + n)|^2 \rangle \\ &= \langle [s - w(Gs + n)][s^* - w^*(G^*s^* + n^*)] \rangle \end{aligned}$$

where the symbol  $\langle \cdot \rangle$  means ensemble average. Then, we have

$$\begin{aligned} \frac{\partial J}{\partial w^*} &= -\langle [s - w(Gs + n)][G^*s^* + n^*] \rangle \\ &= -\langle G^*|s|^2 - w|Gs|^2 - w|n|^2 \rangle \\ &= -G^*A^2 + w|G|^2A^2 + wN \end{aligned}$$

where we assumed  $\langle s^*n \rangle = \langle sn^* \rangle = 0$  and put  $\langle |s|^2 \rangle = A^2$  and  $\langle |n|^2 \rangle = N$ .

Letting  $\frac{\partial J}{\partial w^*} = 0$ , we have

$$w = \frac{G^*A^2}{|G|^2A^2 + N} = \frac{1/G}{1 + N/S}$$

where  $S = |G|^2A^2$  and  $N$  are receive signal and noise powers, respectively. The optimum gain is given by balancing the noise power  $\langle |wn|^2 \rangle$  and squared signal error  $\langle |s - Gws|^2 \rangle$ . We get  $w \rightarrow 1/G$  for  $S/N \rightarrow \infty$ .

(2) *With constraint*

To maximize  $f(x_1, x_2)$  under a constraint as

$$g(x_1, x_2) = C \quad (C : \text{const.})$$

We can take the target function as

$$u(x_1, x_2) = \frac{f(x_1, x_2)}{g(x_1, x_2)} \quad (2.149)$$

where  $u(x_1, x_2)$  is assumed to be differentiable. Then, we find  $\tilde{x}_1, \tilde{x}_2$  to make

$\frac{\partial u}{\partial x_1} = 0, \frac{\partial u}{\partial x_2} = 0$ . By doing partial differentiation of Equation 2.149, we get

$$\begin{aligned}\frac{\partial u}{\partial x_1} &= \frac{1}{g(x_1, x_2)} \frac{\partial}{\partial x_1} f(x_1, x_2) - \frac{f(x_1, x_2)}{(g(x_1, x_2))^2} \frac{\partial}{\partial x_1} g(x_1, x_2) = 0 \\ \frac{\partial u}{\partial x_2} &= \frac{1}{g(x_1, x_2)} \frac{\partial}{\partial x_2} f(x_1, x_2) - \frac{f(x_1, x_2)}{(g(x_1, x_2))^2} \frac{\partial}{\partial x_2} g(x_1, x_2) = 0\end{aligned}$$

The above equations are rewritten as

$$\begin{aligned}\frac{\partial}{\partial x_1} f(x_1, x_2) - \frac{f(x_1, x_2)}{g(x_1, x_2)} \frac{\partial}{\partial x_1} g(x_1, x_2) &= 0 \\ \frac{\partial}{\partial x_2} f(x_1, x_2) - \frac{f(x_1, x_2)}{g(x_1, x_2)} \frac{\partial}{\partial x_2} g(x_1, x_2) &= 0\end{aligned}$$

Denoting  $f(\tilde{x}_1, \tilde{x}_2) = M$ , we get

$$\begin{aligned}\frac{\partial}{\partial x_1} f(x_1, x_2) - \frac{M}{C} \frac{\partial}{\partial x_1} g(x_1, x_2) &= 0 \\ \frac{\partial}{\partial x_2} f(x_1, x_2) - \frac{M}{C} \frac{\partial}{\partial x_2} g(x_1, x_2) &= 0\end{aligned}$$

If we let  $\lambda = M / C$ , the above equations express Lagrange's method of undetermined multiplier. Here, the meaning of the undetermined multiplier is given.

**EXAMPLE 2.28**  $f(x_1, x_2) = \log x_1 + \log x_2$

where

$$x_1 = p_1 s_1, \quad x_2 = p_2 s_2, \quad (2.150)$$

$0 \leq p_1, p_2 \leq 1$ , and  $s_1$  and  $s_2$  are positive constants. As a constraint, we take

$$p_1 + p_2 = 1 \quad (2.151)$$

or

$$g(x_1, x_2) = \frac{x_1}{s_1} + \frac{x_2}{s_2} = 1$$

Using the Lagrange multiplier method, we have

$$\frac{1}{x_1} - \frac{\lambda}{s_1} = 0, \quad \frac{1}{x_2} - \frac{\lambda}{s_2} = 0$$

With the above results, Equations 2.150 and 2.151, we have

$$p_1 = p_2 = 1/2 \quad (2.152)$$

Thus, we get

$$\tilde{x}_1 = \frac{s_1}{2}, \quad \tilde{x}_2 = \frac{s_2}{2} \quad (2.153)$$

The results correspond to the proportional fair scheduling in a resource allocation problem (Section 9.1.3). Without the constraint (Eq. 2.151), we have  $\tilde{x}_1 = s_1, \tilde{x}_2 = 0$  for  $s_1 > s_2$  and  $\tilde{x}_1 = 0, \tilde{x}_2 = s_2$  for  $s_1 < s_2$ .

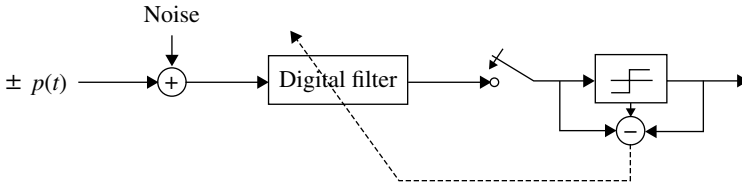
### 2.5.2 Adaptive Signal Processing

Adaptive signal processing means automatic control of variable parameters to make a signal processing system to be optimum adapting to the change of system characteristics. For an example, we can cite automatic control of coefficients of a digital filter (equalizer), which compensates receive signal distortion in a channel. When the channel distortion varies in time, the parameters should be controlled adaptively to the channel distortion. Highly advanced adaptive digital signal processing became feasible due to evolution of semiconductor integrated circuit technologies. In a modern cellular phone, for example, high signal processing, which was not imagined in not so many years ago, is being carried out. In addition to the adaptive digital filter, the other examples include estimation (identification) of an unknown dynamical system and expectation of a varying signal.

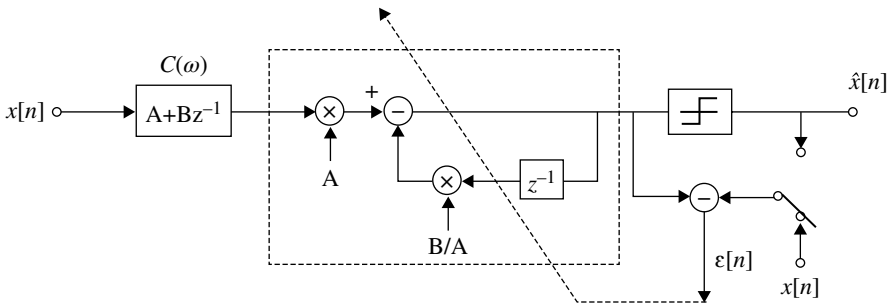
As a measure of performance of this adaptive signal processing, the difference (error) between values of a target and currently obtained one is taken. The signal processing is performed under a criterion to minimize the error. In the following, examples of adaptive signal processing in linear systems are briefly described and the algorithms of adaptive signal processing are explained. Time-discrete (digital) signal processing is assumed.

**Equalization** We consider signal processing (digital filter) at a receiver in a digital communication system as an example (Fig. 2.46). Pulse signals  $\pm p(t)$  are sent at a given time interval  $T$  with its polarity changed to be positive or negative according to the transmit data  $\{1,0\}$  sequence. The received pulse signals are sampled at sampling instants ( $t = nT, n = 0, 1, \dots$ ) and decided to be 1 or 0 depending on the polarities of the sampled pulses. The transmit pulse signal  $\pm p(t)$  suffers from distortion in the channel and the receive signal becomes  $r(t) = \pm p(t) * c(t)$ , where  $c(t)$  denotes the impulse response of the channel.

Let us consider, for example, a multipath radio channel with its impulse response of  $c(t) = A\delta(t) + B\delta(t - t_0)$ , where  $A$  and  $B$  are constants and  $t_0$  is a delay time. Then received signal becomes  $r(t) = \pm\{Ap(t) + Bp(t - t_0)\}$  distorted from the transmit pulse signal  $\pm p(t)$ . Due to the signal distortion and noise added in the channel and/or the receiver, errors in the decision making occur probabilistically. Equalization in a wide



**FIGURE 2.46** Equalization of distorted signal in a digital transmission system.

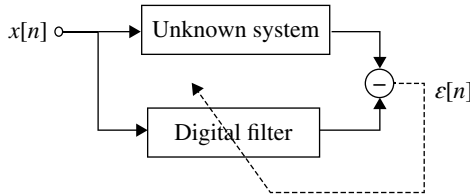
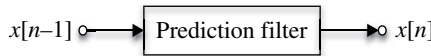
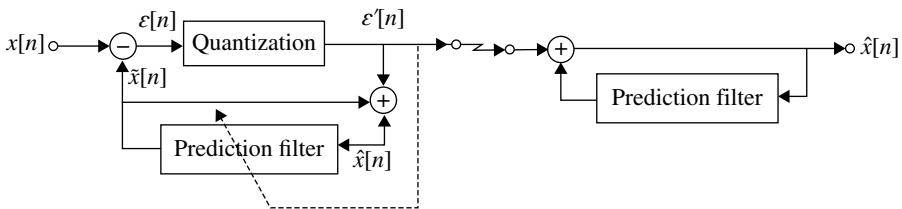


**FIGURE 2.47** An example of equalizer.

sense means to minimize the decision errors. It is sometimes performed by using a digital filter and controlling its coefficients. In a narrow sense, equalization means to recover the transmit pulse waveform from the receive signal without considering the effect of the noise on the decision making.

Here, we consider an equalizer in the narrow sense for simplicity. Then, we should obtain a digital filter as the inverse circuit for the channel with its impulse response  $c(t)$ . In the above channel, the channel impulse is given as  $c(t) = A\delta(t) + B\delta(t - t_0)$ . Taking the Fourier transform of this, we have  $C(\omega) = A + Be^{-j\omega t_0}$ . The inverse circuit has the transfer function as  $H(\omega) = 1/C(\omega) = 1/(A + Be^{-j\omega t_0})$ . If we assume delay time  $t_0$  is equal to the sampling period  $T$  for simplicity, we should get a digital filter with transfer function  $H(z) = 1/(A + Bz^{-1})$ . The circuit is implemented as shown in Figure 2.47. To assure the circuit stability,  $B/A < 1$  is required. Actually, we do not know  $A$ ,  $B$ , and  $t_0$ . Therefore, estimation of them becomes an issue. Since those values change in time, the inverse digital filter should change their coefficients adaptively to follow the change of the channel characteristics. The inverse filter presented here includes a feedback part and may possibly become unstable. Therefore, it is usually implemented as a transversal (or FIR) filter (Fig. 2.40). The number of filter order is decided to balance the circuit complexity and the performance. The control of the filter coefficients is performed to reduce the error signal that is defined as the difference between the equalizer output and the reference level  $\pm p(0)$  of the pulse signal or the known signal sequence  $x[n]$ .

**System Identification** When a system includes some unknown parameters (Fig. 2.48), estimating parameters is called system identification or estimation.

**FIGURE 2.48** System identification.**FIGURE 2.49** Prediction filter.**FIGURE 2.50** An application of prediction filter (ADPCM).

The usage of the estimation tacitly assumes that the parameters are not exactly obtained due to the system fluctuation caused by probabilistic events such as noise. The coefficients of the estimating digital filter is decided to minimize the error defined as the difference between the output  $y[n]$  of the unknown (dynamic) system and  $\hat{y}[n]$  of the estimating filter by applying input signal  $x[n]$  to the unknown system and the estimating filter simultaneously.

*Signal Prediction* We consider to predict a signal at time  $n$  by using signals at times till  $n-1$  (Fig. 2.49). For a linear system, the predicted value  $\hat{x}[n]$  is given by weighting summation of the past values  $x[n-m]$  ( $m = 1, 2, \dots$ ) as

$$\hat{x}[n] = \sum_{m=1}^M h[m]x[n-m] \quad (2.154)$$

This formula is equivalent to a transversal filter output where the coefficient for the present time is set to be zero. The weighting factors  $h[m]$  or the tap coefficients are decided to minimize the prediction error  $\varepsilon[n] = x[n] - \hat{x}[n]$ . When the signal behavior changes in time, the weighting factors should be controlled adaptively to the change.

As an example of the signal prediction, adaptive differential pulse code modulation (ADPCM) of speech signals is shown in Figure 2.50. This voice coding



method can reduce its coding rate to one-half (32 kbps) of the 64 kbps PCM (8 kHz sampling frequency, 8 bits quantization), while keeping the voice quality (quantization error) to be almost the same (Section 7.7.3). The voice signal  $x[n]$  sampled at 8 kHz is subtracted by the prediction filter output  $\hat{x}[n]$  to produce the difference (prediction error)  $\varepsilon[n]$ . The prediction error is quantized ( $\varepsilon'[n]$ ), coded (not shown), and sent to a channel. The quantization error (noise) causes the coded voice quality degradation. A higher number of quantization bits are required to reduce the quantization noise. This in turn results in the increase of the coding rate (speed). Introducing the prediction filter in ADPCM reduces the dynamic range of the prediction error  $\varepsilon[n]$  lower than that of the input voice signal  $x[n]$ , and the number of quantization levels (bits) is reduced. The adaptive control of the prediction filter is carried out based on the quantized signal  $\varepsilon'[n]$  using standardized algorithm. Actually the step size in the quantizer is also adaptively controlled. If we neglect the quantization error, that is,  $\varepsilon[n] = \varepsilon'[n]$ , the input signal to the prediction filter becomes  $\hat{x}[n] = x[n]$ . Then, the receiver (decoder) output signal also becomes equal to the input signal  $x[n]$ .

**Adaptive Optimization Algorithm** We discuss algorithms which control parameters of a linear system to make the system output close to a target signal and reduces the error defined as the difference between the output and the target. Once the error is defined, the algorithm can be applied to any linear system such as a signal equalizer, system identification, and signal prediction systems.

For a discrete-time linear system, the input  $x[n]$  and output  $y[n]$  are related as (Section 2.4.4)

$$y[n] = \sum_{k=0}^{M-1} h[k]x[n-k]$$

where  $h[n]$  ( $n=0, 1, 2, \dots, M-1$ ) denotes the impulse response of the system and the causality of the system is assumed. Denoting a desired (target) signal by  $d[n]$ , the error signal  $e[n]$  is given as

$$e[n] = d[n] - y[n]$$

Input signal  $x[n]$  is assumed to be stochastic, such as random data arrival in digital communications. Even for a deterministic signal, the input signal  $x[n]$  becomes stochastic if noise is added to the signal. The error  $e[n]$  also is stochastic since the input signal is a stochastic variable. As a criterion for adaptive control, minimization of the stochastic average of the error, that is,  $\langle e[n] \rangle$  is not appropriate due to the fact that  $e[n]$  may take a positive or negative large value while keeping  $\langle e[n] \rangle$  to be small through the averaging. Taking the average of absolute value of the error,  $\langle |e[n]| \rangle$ , or of the squared absolute value,  $\langle |e[n]|^2 \rangle$  never shows such an inconvenience. The squared mean error  $\langle |e[n]|^2 \rangle$  is usually used since the optimum parameters are given with a linear operation as follows.

In the following, assuming a real signal system, we discuss some algorithms to find optimum parameters  $h[k](k = 0, 1, \dots, M-1)$  which minimize  $\langle e^2[n] \rangle$  as the criterion. The simplest one applicable to any system is the perturbation method. In this method, starting from given initial values, the parameters are updated step by step to a direction which decreases the target value by changing the parameter values a little in a manner of trial and error around the values in previous step. This method has a disadvantage that the time for converging to the optimum values becomes long due to the trial and error approach. The following improved methods shorten the convergent time.

We rewrite  $\langle e^2[n] \rangle$  as

$$\begin{aligned}\varepsilon &= \langle e^2[n] \rangle = \langle (d[n] - y[n])^2 \rangle \\ &= \left\langle \left( d[n] - \sum_{k=0}^{m-1} h[k]x[n-k] \right)^2 \right\rangle\end{aligned}$$

Since  $\varepsilon$  is a downward-convex function,  $h[k]$  to minimize  $\varepsilon$  always exists. Taking partial differentiation of  $\varepsilon$  with respect to  $h[k]$ , we have

$$\begin{aligned}\frac{\partial \varepsilon}{\partial h[k]} &= \left\langle -2x[n-k] \left( d[n] - \sum_{m=0}^{m-1} h[m]x[n-m] \right) \right\rangle \\ &= -2 \langle x[n-k]e[n] \rangle \quad (k = 0, 1, \dots, M-1)\end{aligned}\tag{2.155}$$

$h[k]$  that satisfy  $\frac{\partial \varepsilon}{\partial h[k]} = 0$  are the optimum parameters, which are denoted as  $h_o[k]$  in the following.  $h_o[k]$  can be obtained by solving the following simultaneous linear equation.

$$\sum_{m=0}^{M-1} \langle x[n-k]x[n-m] \rangle h_o[m] = \langle x[n-k]d[n] \rangle \quad (k = 0, 1, \dots, M-1)\tag{2.156}$$

Here, we assume that  $x[n]$  and  $x[m]$ , and  $x[n]$  and  $d[m]$  are jointly stationary, that is,

$$\langle x[n-k]x[n-m] \rangle = r[m-k]\tag{2.157}$$

$$\langle x[n-k]d[n] \rangle = p[k]\tag{2.158}$$

This means that the joint probabilities never depend on the time origin but on the time difference only.

Using the expression with Equations 2.157 and 2.158, Equation 2.156 is expressed with matrix notation as

$$[R][h_o] = [P]\tag{2.159}$$

where

$$\begin{aligned}
 [h_0] &= \begin{bmatrix} h_0[0] \\ h_0[1] \\ \vdots \\ h_0[M-1] \end{bmatrix} \\
 [P] &= \begin{bmatrix} p[0] \\ p[1] \\ \vdots \\ p[M-1] \end{bmatrix} \\
 [R] &= \begin{bmatrix} r[0] & & r[1] & \cdots & r[M-1] \\ & r[1] & & & \cdots & r[M-2] \\ & \vdots & & & & \\ r[M-1]r[M-2] \cdots & & & & r[0] \end{bmatrix}
 \end{aligned}$$

where we used the property  $r[m-k] = r[k-m]$  of autocorrelation. The optimum parameters  $[h_0]$  are given by solving Equation 2.159 as

$$[h_0] = [R]^{-1}[P]$$

**(a) The Method of Steepest Decent** Finding the optimum coefficients (parameters) by solving simultaneous linear equations or making matrix inversion needs a complex algorithm and a high computation complexity for a large number  $M$  of the coefficients. To solve this difficulty the method of steepest-decent is known. This method is an iterative one where starting from given initial values the optimum values are approached step by step along with the target error surface in a direction to decrease the error value. Denoting explicitly the time  $n$  for  $\varepsilon$  and  $h[k]$ , we express Equation 2.155 as follows:

$$\frac{\partial \varepsilon[n]}{\partial h[k, n]} = -2 \langle x[n-k]e[n] \rangle$$

This equation shows gradient of the error surface and therefore the optimum values lie in a direction to decrease  $h[k, n]$  if the gradient is positive and to increase them if the gradient is negative. Based on this fact, the method is expressed with a recursive equation as

$$\begin{aligned}
 h[k, n+1] &= h[k, n] + \frac{1}{2} \mu \left( -\frac{\partial \varepsilon[n]}{\partial h[k, n]} \right) \\
 &= h[k, n] + \mu \langle x[n-k]e[n] \rangle \quad (k = 0, 1, \dots, M-1)
 \end{aligned} \tag{2.160}$$

where  $\mu$  called a step size parameter is a positive small constant. When we get  $\langle x[n-k]e[n] \rangle = 0$ , the optimum coefficients are reached and the update stops.

**(b) Least Mean Square Algorithm** The steepest-decent method needs to calculate  $\langle x[n-k]e[n] \rangle$  which corresponds to the gradients, where the stochastic average should be taken. Taking the stochastic average in real time becomes difficult since it requires many input data samples. The least mean square (LMS) algorithm substitutes an instant value in place of the stochastic average as

$$h[k, n+1] = h[k, n] + \mu x[n-k]e[n] \quad (k = 0, 1, \dots, M-1)$$

Thus, we can update the coefficients at every time when a new input data  $x[n]$  is applied. The value of  $\mu$  affects the convergent speed and the residual error: A large  $\mu$  yields a high residual error, although the convergent time becomes short. The situation reverses for a small  $\mu$ .

The level of input signal  $x[n]$  has also a similar effect as seen from Equation 2.160. To prevent the effect in LMS algorithm, automatic gain control of the input signal is introduced. An equivalent technique called the normalized LMS is known as

$$h[k, n+1] = h[k, n] + \frac{\alpha}{\|x\|^2} x[n-k]e[n] \quad (k = 0, 1, \dots, M-1)$$

where  $\|x\|^2 = \sum_{k=0}^{M-1} (x[n-k])^2$  and  $\alpha$  is a positive small constant.

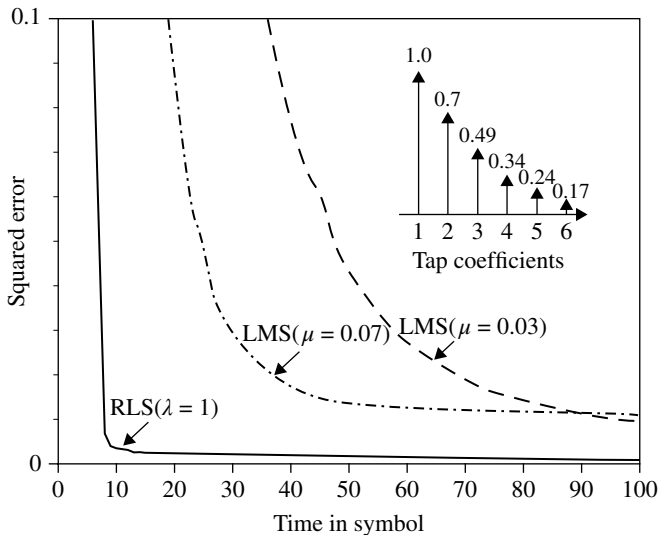
**(c) Recursive Least Square Algorithm** First, examples of convergent characteristics with LMS and recursive least square (RLS) algorithm are shown in Figure 2.51, when an inverse filter against a distorting channel with a transfer function  $H(z) = 1 + z^{-2}/2$  is implemented by a transversal filter. The convergent speed of the LMS algorithm is slow, since the update of coefficients is carried out sequentially on instantaneous errors. The RLS algorithm obtains the optimum coefficients by using all the input data applied so far and shows a higher convergent speed as shown in the figure. This method is also based on an iterative manner. The computational complexity for the method is, however, higher than that for the LMS method.

Let us assume that an input (measured) data sequence of  $n$  elements,  $x[i]$  ( $i = 1, 2, \dots, n$ ) is given. A target or desired signal data sequence corresponding to the input data is denoted as  $d[i]$  ( $i = 1, 2, \dots, n$ ). We will obtain for a linear digital filter the  $M$  optimum coefficients  $h[k, n]$  ( $m = 1, 2, \dots, M-1$ ) which minimize the squared value  $J[n]$  of an error defined as the difference between the desired data sequence and the output of the filter when the input data sequence is applied. The index  $n$  is used to explicitly show that  $n$  data are under consideration. The squared error  $J[n]$  is given as

$$J[n] = \sum_{i=1}^n e^2[i] \quad (2.161)$$

where

$$e[i] = d[i] - y[i] \quad (2.162)$$



**FIGURE 2.51** Examples of convergent characteristics for LMS and RLS.

$$y[i] = \sum_{k=0}^{M-1} h[k, n] x[i - k] \quad (i = 1, 2, \dots, n) \quad (2.163)$$

We assume  $x[i - k] = 0$ , if  $i - k < 1$ .

Inserting Equations 2.162 and 2.163 into Equation 2.161, we have

$$J[n] = \sum_{i=1}^n \left( d[i] - \sum_{k=0}^{M-1} h[k, n] x[i - k] \right)^2$$

Taking partial differentiation with respect to  $h[k, n]$  yields

$$\frac{\partial J[n]}{\partial h[k, n]} = -2 \sum_{i=1}^n x[i - k] \left( d[i] - \sum_{m=0}^{M-1} h[m, n] x[i - m] \right)$$

Denoting by  $\hat{h}[k, n]$  for  $h[k, n]$  which satisfies  $\partial J[n] / \partial h[k, n] = 0$  and, therefore, minimizes  $J[n]$ , we have

$$\sum_{m=0}^{M-1} \hat{h}[m, n] \sum_{i=1}^n x[i - k] x[i - m] = \sum_{i=1}^n x[i - k] d[i] \quad (k = 0, 1, \dots, M-1) \quad (2.164)$$

The optimum coefficients  $\hat{h}[m, n]$  ( $m = 0, 1, \dots, M-1$ ) are given by solving the above simultaneous linear equations. Those equations are similar to Equation 2.156. The difference between them exists in the fact that auto- and cross-correlations are defined as ensemble averages in Equation 2.156 while they are defined by time averages in Equation 2.164. For simplicity, we denote time (average) correlations as

$$\begin{aligned}
\phi[n; k, m] &= \sum_{i=1}^n x[i-k]x[i-m] \quad (k, m = 0, 1, \dots, M-1) \\
\theta[n; k] &= \sum_{i=1}^n d[i]x[i-k] \quad (k = 0, 1, \dots, M-1)
\end{aligned} \tag{2.165}$$

From this, Equation 2.164 becomes

$$\sum_{m=0}^{M-1} \phi[n; k, m] \hat{h}[m, n] = \theta[n; k] \quad (k = 0, 1, \dots, M-1)$$

We express them with a matrix equation as

$$[\phi[n]] [\hat{h}[n]] = [\theta[n]] \tag{2.166}$$

where  $[\phi[n]]$  is a matrix  $(M \times M)$  with its elements  $\phi[n; k, m]$  and  $[\hat{h}[n]]$ , and  $[\theta[n]]$  are column vectors with  $M$  elements or matrices  $(M \times 1)$  as

$$[\hat{h}[n]] = \begin{bmatrix} \hat{h}[0, n] \\ \hat{h}[1, n] \\ \vdots \\ \hat{h}[M-1, n] \end{bmatrix}, \quad [\theta[n]] = \begin{bmatrix} \theta[n; 0] \\ \theta[n; 1] \\ \vdots \\ \theta[n; M-1] \end{bmatrix}$$

The matrix  $[\phi[n]]$  becomes symmetrical due to the autocorrelation property ( $\phi[n; k, m] = \phi[n; m, k]$ ). Solving Equation 2.166, we get the optimum solution  $[\hat{h}[n]]$  as

$$[\hat{h}[n]] = [\phi[n]]^{-1} [\theta[n]] \tag{2.167}$$

The RLS algorithm is a method to solve the above equation iteratively. As a preparation to proceed, we show a method to iteratively calculate  $\phi[n; k, m]$  and  $\theta[n; k]$ . We rewrite Equation 2.165 as

$$\phi[n; k, m] = \sum_{i=1}^{n-1} x[i-k]x[i-m] + x[n-k]x[n-m]$$

The first term in the right side shows  $\phi[n-1; k, m]$ . Therefore, we have an iterative equation as

$$\phi[n; k, m] = \phi[n-1; k, m] + x[n-k]x[n-m] \quad (m, k = 0, 1, \dots, M-1) \tag{2.168}$$

Similarly, we get

$$\theta[n; k] = \theta[n-1; k] + d[n]x[n-k] \quad (k = 0, 1, \dots, M-1). \tag{2.169}$$

Since we use many matrix expression in the following, we express a matrix with a block letter for simplicity as  $[\phi[n]] = \phi[n]$  for example. We define a vector of  $M$  input data from the past to the present instant  $n$  as

$$\mathbf{x}[n] = \begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-M+1] \end{bmatrix}$$

Then Equation 2.168 is expressed with matrices as

$$\phi[n] = \phi[n-1] + \mathbf{x}[n]\mathbf{x}^T[n] \quad (\mathbf{x}^T : \text{transpose of } x) \quad (2.170)$$

Similarly, Equation 2.169 is expressed with column vectors as

$$\theta[n] = \theta[n-1] + d[n]\mathbf{x}[n] \quad (2.171)$$

Equation 2.167 is rewritten as

$$\hat{\mathbf{h}}[n] = \phi[n]^{-1}\theta[n] \quad (2.172)$$

The matrix-inversion lemma is known as follows: If a matrix  $\mathbf{A}$  is given as

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^T$$

where  $\mathbf{A}$  and  $\mathbf{B}$  are  $M \times M$  matrices,  $\mathbf{C}$  is an  $M \times N$  matrix, and  $\mathbf{D}$  is an  $N \times N$  matrix, then the inverse matrix of  $\mathbf{A}$  is given as.

$$\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B}\mathbf{C}[\mathbf{D} + \mathbf{C}^T\mathbf{B}\mathbf{C}]^{-1}\mathbf{C}^T\mathbf{B}$$

To apply the lemma to Equation 2.170, we let  $\mathbf{A} = \phi[n]$ ,  $\mathbf{B}^{-1} = \phi[n-1]$ ,  $\mathbf{C} = \mathbf{x}[n]$ ,  $\mathbf{D} = \mathbf{I}$  (identity matrix). Then, we have

$$\phi^{-1}[n] = \phi^{-1}[n-1] - \frac{\phi^{-1}[n-1]\mathbf{x}[n]\mathbf{x}^T[n]\phi^{-1}[n-1]}{1 + \mathbf{x}^T[n]\phi^{-1}[n-1]\mathbf{x}[n]} \quad (2.173)$$

To use simplified expression, we let

$$\mathbf{P}[n] = \phi^{-1}[n] \quad (2.174)$$

$$\mathbf{k}[n] = \frac{\mathbf{P}[n-1]\mathbf{x}[n]}{1 + \mathbf{x}^T[n]\mathbf{P}[n-1]\mathbf{x}[n]} \quad (2.175)$$

then Equation 2.173 is written as

$$\mathbf{P}[n] = \mathbf{P}[n-1] - \mathbf{k}[n]\mathbf{x}^T[n]\mathbf{P}[n-1] \quad (2.176)$$

Multiplying  $\mathbf{x}[n]$  to the both sides of the equation from the right side, we get

$$\mathbf{P}[n]\mathbf{x}[n] = \mathbf{P}[n-1]\mathbf{x}[n] - \mathbf{k}[n]\mathbf{x}^T[n]\mathbf{P}[n-1]\mathbf{x}[n] \quad (2.177)$$

We rewrite Equation 2.175 as

$$\mathbf{k}[n] = \mathbf{P}[n-1]\mathbf{x}[n] - \mathbf{k}[n]\mathbf{x}^T[n]\mathbf{P}[n-1]\mathbf{x}[n] \quad (2.178)$$

From Equations 2.177 and 2.178, we have

$$\mathbf{k}[n] = \mathbf{P}[n]\mathbf{x}[n] \quad (2.179)$$

Inserting Equations 2.171, 2.174, and 2.179 into Equation 2.172, we get

$$\begin{aligned} \hat{\mathbf{h}}[n] &= \mathbf{P}[n]\boldsymbol{\theta}[n] \\ &= \mathbf{P}[n]\{\boldsymbol{\theta}[n-1] + \mathbf{x}[n]d[n]\} \\ &= \mathbf{P}[n]\boldsymbol{\theta}[n-1] + \mathbf{P}[n]\mathbf{x}[n]d[n] \\ &= \mathbf{P}[n]\boldsymbol{\theta}[n-1] + \mathbf{k}[n]d[n] \end{aligned}$$

Inserting  $\mathbf{P}[n]$  (Eq. 2.176), we have

$$\begin{aligned} \hat{\mathbf{h}}[n] &= \{\mathbf{P}[n-1] - \mathbf{k}[n]\mathbf{x}^T[n]\mathbf{P}[n-1]\}\boldsymbol{\theta}[n-1] + \mathbf{k}[n]d[n] \\ &= \mathbf{P}[n-1]\boldsymbol{\theta}[n-1] + \mathbf{k}[n]\{d[n] - \mathbf{x}^T[n]\mathbf{P}[n-1]\boldsymbol{\theta}[n-1]\} \end{aligned} \quad (2.180)$$

From Equations 2.172 and 2.174, we obtain

$$\mathbf{P}[n-1]\boldsymbol{\theta}[n-1] = \hat{\mathbf{h}}[n-1]$$

and Equation 2.180 becomes

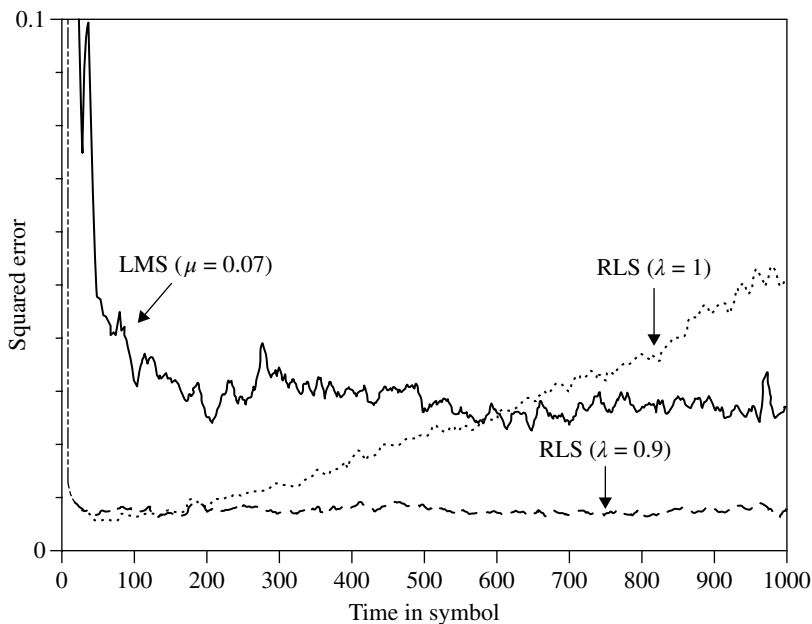
$$\begin{aligned} \hat{\mathbf{h}}[n] &= \hat{\mathbf{h}}[n-1] + \mathbf{k}[n]\{d[n] - \mathbf{x}^T[n]\hat{\mathbf{h}}[n-1]\} \\ &= \hat{\mathbf{h}}[n-1] + \mathbf{k}[n]\eta[n] \end{aligned} \quad (2.181)$$

where

$$\eta[n] = d[n] - \mathbf{x}^T[n]\hat{\mathbf{h}}[n-1] \quad (2.182)$$

$\eta[n]$  means “error” when the coefficients at time  $n-1$  are used. Thus,  $\hat{\mathbf{h}}[n]$  can be obtained recursively with Equations 2.175, 2.176, 2.181, and 2.182 if the initial values are given. We put all the initial values to be zero ( $\hat{\mathbf{h}}[0] = \mathbf{0}$ ). If we put  $\phi[0; k, m] = 0$  (Eq. 2.165),  $\mathbf{P}[0]$  becomes infinity. Therefore, we assume  $\phi[0] = c\mathbf{I}$ ,





**FIGURE 2.52** Convergent characteristics for a nonstationary system.

where  $c$  is a positive small constant and  $\mathbf{I}$  is an  $M \times M$  identity matrix. When  $c$  comes closer to zero, we can confirm an  $\hat{\mathbf{h}}[n]$  approach to solutions, which is given by directly solving Equation 2.167.

It is shown from the above arguments that RLS algorithm gives coefficients  $\hat{\mathbf{h}}[n]$  to minimize the mean square error when  $n$  input  $x[i]$  data and desired data  $d[i]$  ( $i = 0, 1, \dots, n$ ) are given.

So far we assumed a stationary system. If the system is not stationary, the RLS algorithm yields a high error. The situation is shown in Figure 2.52. To apply the RLS algorithm much better to a nonstationary system, we modify the mean square error as

$$J[n] = \sum_{i=1}^N \lambda^{n-i} e^2[i]$$

where  $\lambda$  is a constant that takes a value  $0 < \lambda \leq 1$ . By taking  $\lambda < 1$ , the effect of the past data is gradually lost (forgotten) and the coefficients are controlled adaptively to the change of the system. By modifying the error like the above, we have

$$\begin{aligned} \phi[n; k, m] &= \lambda \phi[n-1; k, m] + x[n-k]x[n-m] \\ \theta[n; k] &= \lambda \theta[n-1; k] + d[n]x[n-k] \end{aligned}$$

Thus, we should replace variables as  $\phi[n] \rightarrow \lambda\phi[n]$  and therefore  $P[n-1] \rightarrow \lambda^{-1}P[n-1]$  and  $\theta[n-1] \rightarrow \lambda\theta[n-1]$  in the former arguments. The successive equations are summarized as follows:

Initial conditions:  $P[0] = c^{-1}I$ ,  $\hat{h}[0] = \mathbf{0}$

$$\begin{aligned} k[n] &= \frac{\lambda^{-1}P[n-1]\mathbf{x}[n]}{1 + \lambda^{-1}\mathbf{x}^T[n]P[n-1]\mathbf{x}[n]} \\ \eta[n] &= d[n] - \mathbf{x}^T[n]\hat{h}[n-1] \\ \hat{h}[n] &= \hat{h}[n-1] + k[n]\eta[n] \\ P[n] &= \lambda^{-1}P[n-1] - \lambda^{-1}k[n]\mathbf{x}^T[n]P[n-1] \end{aligned}$$

## APPENDIX 2.A $\lim_{\Omega \rightarrow \infty} (\sin \Omega t / \pi t) = \delta(t)$

We will show the following equation holds.

$$\lim_{\Omega \rightarrow \infty} \int_{-\infty}^{\infty} f(t) \frac{\sin \Omega t}{\pi t} dt = f(0) \quad (2.A.1)$$

The integral is divided into three regions as follows:

$$\int_{-\infty}^{-\varepsilon} + \int_{-\varepsilon}^{\varepsilon} + \int_{\varepsilon}^{\infty} f(t) \frac{\sin \Omega t}{\pi t} dt \quad (0 < \varepsilon < 1)$$

The second integral becomes

$$\lim_{\Omega \rightarrow \infty} \int_{-\varepsilon}^{\varepsilon} f(t) \frac{\sin \Omega t}{\pi t} dt = f(0) \lim_{\Omega \rightarrow \infty} \int_{-\varepsilon}^{\varepsilon} \frac{\sin \Omega t}{\pi t} dt = f(0) \quad (\varepsilon \rightarrow 0)$$

where we used  $x = \Omega t$  and

$$\lim_{\Omega \rightarrow \infty} \int_{-\varepsilon}^{\varepsilon} \frac{\sin \Omega t}{\pi t} dt = \lim_{\Omega \rightarrow \infty} \int_{-\Omega\varepsilon}^{\Omega\varepsilon} \frac{\sin x}{\pi x} dx = 1$$

where  $\int_{-\infty}^{\infty} \sin x / x dx = \pi$  is used.

If  $f(t)$  is discontinuous at  $t = 0$ , since  $\int_{-\infty}^0 \sin x / x dx = \pi / 2$ , the second integral becomes as  $[f(0)^+ + f(0)^-] / 2$  where  $f(0^+)$  and  $f(0^-)$  are right-hand and left-hand limits, respectively.

We rewrite the third integral as  $\lim_{T, \Omega \rightarrow \infty} \int_0^T g(t) \sin \Omega t dt$ , where  $g(t) = f(t) / \pi t$ . This integral becomes zero if  $f(t)$  and, therefore,  $g(t)$  satisfies the conditions given in Appendix 2.B, where we assume  $\varepsilon > 0$ . Similarly, the first integral also becomes zero.

Let us consider  $\cos \Omega t$  instead of  $\sin \Omega t$  in the above arguments. Then, we can show that  $\lim_{\Omega \rightarrow \infty} \int_{-\infty}^{\infty} f(t) \cos \frac{\Omega t}{\pi t} dt = 0$ . Thus, we get the following expression:

$$\lim_{\Omega \rightarrow \infty} \frac{e^{j\Omega t}}{j\pi t} = \delta(t) \quad (2.A.2)$$

## APPENDIX 2.B CONDITIONS FOR A TEST FUNCTION FOR THE DELTA FUNCTION, $\lim_{T, \Omega \rightarrow \infty} \int_{\varepsilon}^T g(t) \sin \Omega t dt = 0$

A condition  $\int_{\varepsilon}^T |g(t)| dt < \infty$  to fulfill the above equation is known as the Riemann–Lévesque theorem. Here, we will show other sufficient conditions. Since  $e^{j\omega t} = \cos \omega t + j \sin \omega t$ , it is sufficient to show that  $\lim_{T, \Omega \rightarrow \infty} \int_{\varepsilon}^T g(t) e^{j\Omega t} dt = 0$ . We integrate by part as

$$\int_{\varepsilon}^T g(t) e^{j\Omega t} dt = \frac{1}{j\Omega} \left[ g(t) e^{j\Omega t} \right]_{\varepsilon}^T - \frac{1}{j\Omega} \int_{\varepsilon}^T g'(t) e^{j\Omega t} dt$$

If  $g(t)$  is bounded (condition (i)), then the first term in the right-hand side becomes zero when  $\Omega \rightarrow \infty$ . Since

$$\int_{\varepsilon}^T g'(t) e^{j\Omega t} dt \leq \int_{\varepsilon}^T |g'(t)| dt$$

The second term becomes zero for  $\Omega \rightarrow \infty$ , if  $\int_{\varepsilon}^T |g'(t)| dt < \infty$  ( $\Omega \rightarrow \infty$ ), that is,  $g'(t)$  is absolutely integrable (condition (ii)).

If  $|g'(t)|$  is bounded (less than  $M$ ) (condition (iii)), the second term becomes zero for a finite  $T$ , since

$$\int_{\varepsilon}^T g'(t) e^{j\Omega t} dt \leq \int_{\varepsilon}^T M dt = M(T - \varepsilon)$$

Even when  $T = \infty$ , if we restrict that  $\lim_{T, \Omega \rightarrow \infty} T/\Omega = 0$  for  $T \rightarrow \infty$  and  $\Omega \rightarrow \infty$ , then the second term becomes zero, since the second term is upper-limited as  $\lim_{T, \Omega \rightarrow \infty} M(T - \varepsilon)/\Omega$ .

The above argument holds for  $\varepsilon = -T$ . Therefore,

$$\lim_{\Omega \rightarrow \pm\infty} \int_{-T}^T g(t) e^{j\Omega t} dt = 0 \quad (2.B.1)$$

In a sense of such a generalized extreme, we get

$$\lim_{\Omega \rightarrow \pm\infty} e^{j\Omega t} = \lim_{\Omega \rightarrow \pm\infty} \cos \Omega t = \lim_{\Omega \rightarrow \pm\infty} \sin \Omega t = 0 \quad (2.B.2)$$

Similarly, in a sense that,  $\lim_{T \rightarrow \pm\infty} \int_{-\Omega}^{\Omega} G(\omega) e^{j\omega T} d\omega = 0$ , the following expression is given as

$$\lim_{T \rightarrow \pm\infty} e^{j\omega T} = \lim_{T \rightarrow \pm\infty} \cos \omega T = \lim_{T \rightarrow \pm\infty} \sin \omega T = 0 \quad (2.B.3)$$

An example of functions that never satisfy the conditions (i)–(iii) is given as  $g(t) = 1/t$  when the integrate region includes 0. In this case,

$$\begin{aligned} \lim_{\Omega \rightarrow \pm\infty} \frac{e^{j\Omega t}}{t} &= \pm j\pi\delta(t), \\ \lim_{\Omega \rightarrow \pm\infty} \sin \frac{\Omega t}{t} &= \pm\pi\delta(t), \\ \lim_{\Omega \rightarrow \pm\infty} \cos \frac{\Omega t}{t} &= 0. \end{aligned}$$

Equation 2.B.1 expresses the Fourier integral  $G(\omega) \leftrightarrow g(t)$  and  $G(\pm\infty) = 0$ , that is, frequency components at infinite frequencies becomes zero.

## APPENDIX 2.C FORMULAE FOR THE TRIGONOMETRIC FUNCTIONS

$$\begin{aligned} \cos(x+y) &= \cos x \cos y - \sin x \sin y \\ \sin(x+y) &= \sin x \cos y + \cos x \sin y \\ 2 \cos x \cos y &= \cos(x+y) + \cos(x-y) \\ 2 \cos x \sin y &= \sin(x+y) - \sin(x-y) \\ 2 \sin x \cos y &= \sin(x+y) + \sin(x-y) \\ 2 \sin x \sin y &= -\cos(x+y) + \cos(x-y) \\ 2 \cos^2 x &= 1 + \cos 2x \\ 2 \sin^2 x &= 1 - \cos 2x \end{aligned}$$

## REFERENCES

- [1] Dirac PAM. *The Principles of Quantum Mechanics*. Oxford: Oxford University Press; 1958.
- [2] Papoulis A. *The Fourier Integral and Its Application*. New York: McGraw-Hill; 1962.
- [3] Harmuth HF. *Transmission of Information by Orthogonal Functions*. New York: Springer-Verlag; 1970.
- [4] Schetzen M. *The Volterra and Wiener Theories of Nonlinear Systems*. Melbourne: Krieger Publishing Company; 2006.
- [5] Lathi BP. *Modern Digital and Analog Communication Systems*. New York: Holt, Rinehart and Winston; 1983.

---

# 3

---

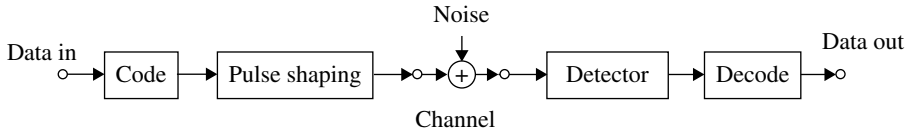
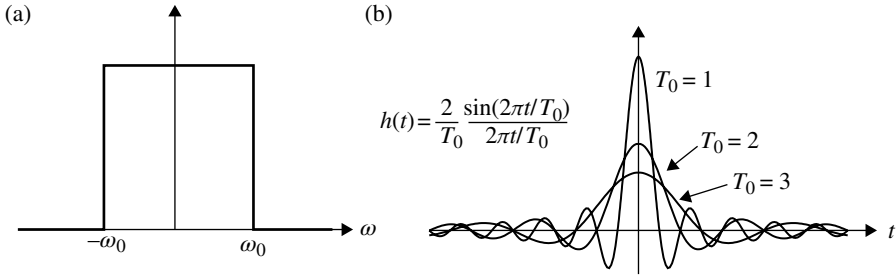
## THE ELEMENTS OF DIGITAL COMMUNICATION SYSTEM

This chapter briefly covers some general topics on digital signal transmission and detection theory. The reader may refer Refs. [1–10] for more comprehensive details.

Figure 3.1 shows the block diagram of a baseband digital transmission system. The transmitted binary data symbols are applied at the pulse repetition frequency to an encoding circuit that performs logical manipulations such as differential encoding and two-to-multilevel conversion. Corresponding to the logical data symbols, the pulse-shaping circuit sends out digitally encoded waveforms into the channel. Pulse shaping is intended to get a narrower signal spectrum or to match the spectrum to the transfer characteristics of the channel, for example, a dc blocked channel. The manipulation of data by the encoder or pulse-shaping circuits is called *line coding*. The transmitted signal is detected at the receiver. Typically, the transmitted signal is corrupted by noise and interference in the channel. The detection process includes noise band limitation and decision-making on the digital signal. Achieving a lower error rate for the decisions is the main concern in the receiver. The determined logical symbol is decoded appropriately corresponding to the encoding.

### 3.1 PULSE SHAPING

Pulse shaping with limited bandwidth results in an impulse response of the pulse-shaping circuit/filter with a long time period. For example, let us consider an ideal low-pass filter with a rectangular transfer function  $H(\omega)$  as shown in Figure 3.2a. The impulse response is given by the inverse Fourier transform as

**FIGURE 3.1** Digital transmission system.**FIGURE 3.2** Characteristics of an ideal low-pass filter.  $T_0 = 2\pi/\omega_0$ . (a) Transfer function and (b) impulse response.

$$\begin{aligned}
 h(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) e^{j\omega t} d\omega \\
 &= \frac{\omega_0 \sin(\omega_0 t)}{\pi \omega_0 t} = \frac{2 \sin(2\pi t/T_0)}{T_0 2\pi t/T_0}
 \end{aligned} \tag{3.1}$$

where  $T_0 = 2\pi/\omega_0$ .

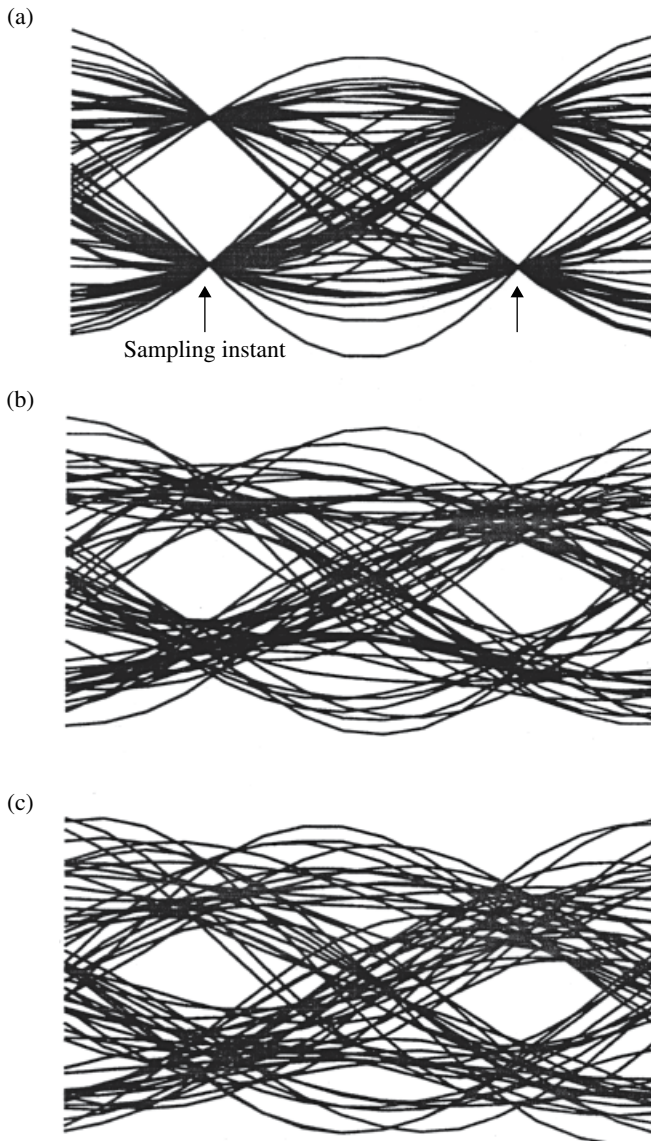
$h(t)$  is drawn in Figure 3.2b. The part of  $h(t)$  that spreads into other symbol periods causes interference with signals, namely, intersymbol interference. The intersymbol interference degrades decision error rate performance at the receiver. Figure 3.3, known as eye diagram, shows an overlapped drawing over two symbol periods of a randomly generated binary data sequence with pulse shaping of Equation 3.1. At every symbol time, the signal is sampled and subjected to decision-making. The minimum distance between the sampled signal levels corresponding to different binary signals is called eye opening. The ideal eye opening is given when the intersymbol interference is null at the sampling instant, as shown in Figure 3.3a.

### 3.1.1 Nyquist's First Criterion

Nyquist's first criterion guarantees that no intersymbol interference occurs. It requires that the impulse response  $h(t)$  of a pulse-shaping filter satisfies the following:

$$h(t) = \begin{cases} 1 & (t = 0) \\ 0 & (t = nT, n \neq 0) \end{cases} \tag{3.2}$$

where  $1/T$  is the pulse repetition frequency. Decisions is made at  $t=nT$ .



**FIGURE 3.3** Eye diagrams of different pulse shaping.  $T$  is pulse duration. (a)  $T/T_0=0.5$ , (b)  $T/T_0=0.475$ , and (c)  $T/T_0=0.45$ .

Considering the above requirement for no intersymbol interference, the sampled impulse response  $h_s(t)$  is given as:

$$h_s(t) = h(t) \sum_{n=-\infty}^{\infty} h(t-nT) = \delta(t)$$

Let  $h(t) \leftrightarrow H(\omega)$ . Taking the Fourier transform of the above equation, we obtain, in similar process to Equation 2.108,

$$\frac{1}{T} \sum_{n=-\infty}^{\infty} H(\omega - n\omega_0) = 1$$

or

$$\sum_{n=-\infty}^{\infty} H(\omega - n\omega_0) = T \quad (3.3)$$

where  $\omega_0 = 2\pi/T$ .

Let us assume  $H(\omega)$  is band-limited within  $0 < |\omega| < \omega_0$ . If we consider a frequency range  $0 \sim \omega_0$ , then Equation 3.3 becomes

$$H(\omega) + H(\omega - \omega_0) = T$$

Consider a new variable  $x = \omega - \omega_0/2$ , then

$$H(\omega_0/2 + x) + H(x - \omega_0/2) = T \left( |x| < \frac{\omega_0}{2} \right)$$

Applying Equation 2.37 to the above equation, we have,

$$H\left(\frac{\omega_0}{2} + x\right) + H^*\left(\frac{\omega_0}{2} - x\right) = T \quad \left( |x| < \frac{\omega_0}{2} \right) \quad (3.4)$$

Taking the real part and the imaginary part of Equation 3.4,

$$\text{Re} \left\{ H\left(\frac{\omega_0}{2} + x\right) \right\} + \text{Re} \left\{ H\left(\frac{\omega_0}{2} - x\right) \right\} = T \quad \left( |x| < \frac{\omega_0}{2} \right) \quad (3.5)$$

$$\text{Im} \left\{ H\left(\frac{\omega_0}{2} + x\right) \right\} - \text{Im} \left\{ H\left(\frac{\omega_0}{2} - x\right) \right\} = 0 \quad \left( |x| < \frac{\omega_0}{2} \right) \quad (3.6)$$

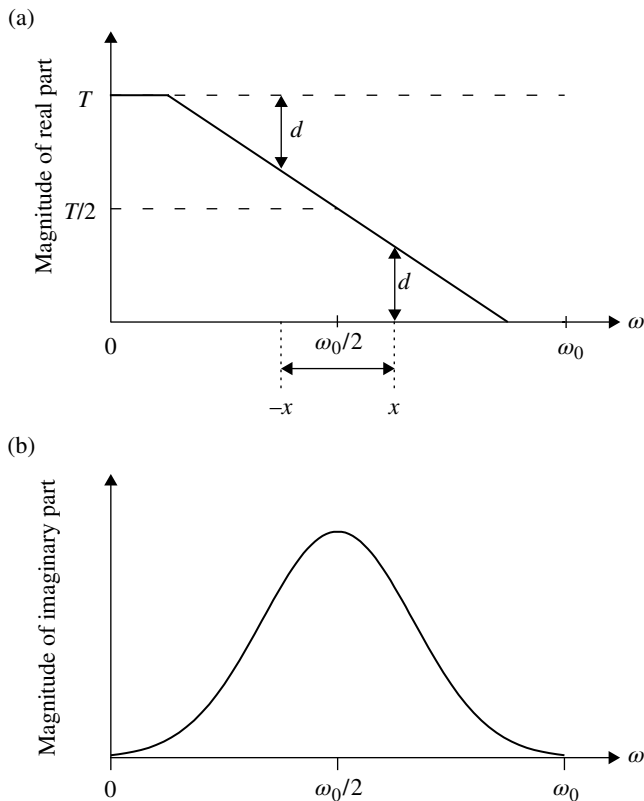
These relations are illustrated in Figure 3.4. There are a number of functions to fulfill the above conditions. If we consider a real function  $G(\omega)$ , then Equation 3.6 is automatically satisfied. However, a transfer function that has no imaginary part could not be realized.

Thus, the realizable transfer function can be in the following form:

$$H(\omega) = G(\omega) e^{-j\omega t_0} \quad (3.7)$$

where  $G(\omega)$  is a real function to meet Equation 3.5 and  $t_0$  is a time delay constant. Because the term  $e^{-j\omega t_0}$  represents a transfer function of a delay line, it does not distort the transmitted waveform. Since  $G(\omega)$  is a real function, the impulse response ( $g(t) \leftrightarrow G(\omega)$ ) becomes an even function, that is,  $g(-t) = g(t)$ . Then from Equations 3.7 and 2.29, we have  $h(t) = g(t - t_0)$ . Since  $G(\omega)$  is band-limited within the finite





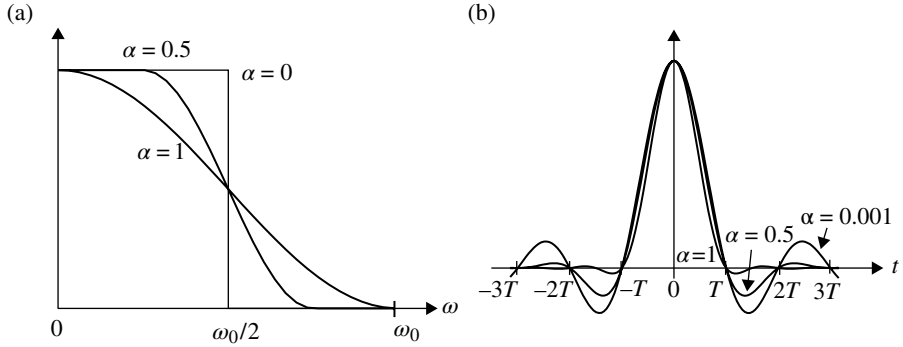
**FIGURE 3.4** Transfer characteristics to meet Nyquist's first criterion. (a) Real part of the transfer function and (b) imaginary part of the transfer function.

bandwidth  $|\omega| < \omega_0$ ,  $g(t)$  shows an infinite response (Section 2.1.2). Therefore, the time delay  $t_0$  must be infinite to get an ideal Nyquist's-I transfer function. Practically, approximations are made for reasonable values of  $t_0$ .

A widely used filter to meet Nyquist's first criterion is the raised cosine roll-off filter, whose transfer function is given as:

$$G(\omega) = \begin{cases} 1 & \left( |\omega| < \frac{(1-\alpha)\omega_0}{2} \right) \\ \frac{1}{2} \left\{ 1 + \cos \left[ \left( \frac{|\omega| - \frac{\omega_0}{2}}{\frac{\alpha\omega_0}{2}} + 1 \right) \frac{\pi}{2} \right] \right\} & \left( \frac{(1-\alpha)\omega_0}{2} \leq |\omega| \leq \frac{(1+\alpha)\omega_0}{2} \right) \\ 0 & \left( |\omega| > \frac{(1+\alpha)\omega_0}{2} \right) \end{cases} \quad (3.8)$$

where  $\alpha (\leq 1)$  is a roll-off factor.



**FIGURE 3.5** (a) Transfer characteristics and (b) impulse response of a raised cosine roll-off filter for different roll-off factors.

The impulse response of the filter is given by

$$g(t) = \frac{1}{T} \frac{\sin \pi(t/T)}{\pi(t/T)} \frac{\cos \alpha \pi(t/T)}{1 - 4\alpha^2 t^2 / T^2} \quad (3.9)$$

Figures 3.5 and 3.6 show  $G(\omega)$ ,  $h(t)$ , and eye diagrams of the raised cosine roll-off filter.

### 3.1.2 Nyquist's Second Criterion

This criterion guarantees an eye opening at the midpoint of symbol times. It requires an impulse response and is as follows:

$$h(t) = \begin{cases} 1 & \left( t = \pm \frac{T}{2} \right) & (n = 0) \\ 0 & \left( t = \pm \left( n + \frac{1}{2} \right) T \right) & (n \neq 0) \end{cases}$$

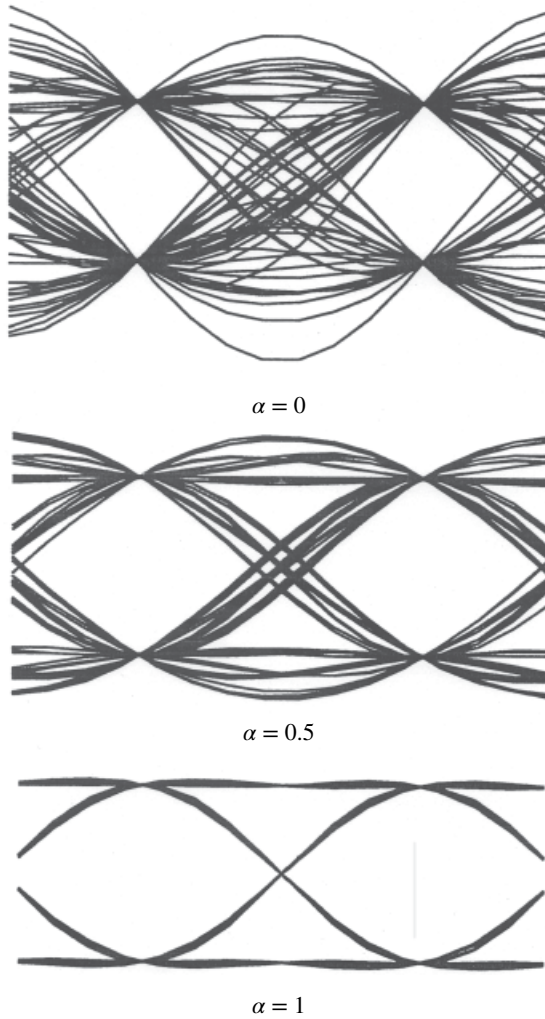
Multiplying the sampling function with the above, we get:

$$h(t) \sum_{n=-\infty}^{\infty} \delta[t - (n + 1/2)T] = \delta(t + T/2) + \delta(t - T/2)$$

Taking the Fourier transform, we have:

$$\sum_{n=-\infty}^{\infty} H(\omega - n\omega_0) e^{-jn\pi} = 2T \cos \frac{\omega T}{2}$$

where  $\omega_0 = 2\pi/T$ .

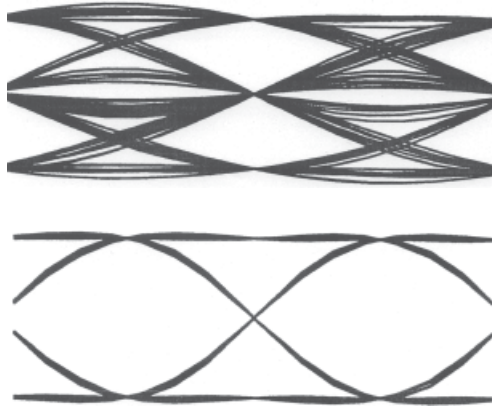


**FIGURE 3.6** Eye diagram of a raised cosine roll-off filter for different roll-off factors.

If we restrict the bandwidth of  $H(\omega)$  within a frequency range of  $|\omega| \leq \omega_0/2$ , then from the above equation we have:

$$H(\omega) = \begin{cases} 2T \cos\left(\frac{\omega T}{2}\right) & \left(|\omega| \leq \frac{\omega_0}{2}\right) \\ 0 & (\text{otherwise}) \end{cases} \quad (3.10)$$

If we restrict the bandwidth of  $H(\omega)$  within  $|\omega| \leq \omega_0$  and only consider the frequency over  $0 < \omega < \omega_0$ , then we get:



**FIGURE 3.7** Eye diagrams of pulse shaping with Nyquist's second criterion. (Top)  $H(\omega)$ : Equation 3.10; (bottom)  $H(\omega)$ : Equation 3.11.

$$H(\omega) - H(\omega - \omega_0) = 2T \cos\left(\frac{\omega T}{2}\right) \quad (0 < \omega < \omega_0)$$

The following transfer function satisfies the above condition:

$$H(\omega) = \begin{cases} T \left[ 1 + \cos\left(\frac{\omega T}{2}\right) \right] & (0 < |\omega| < \omega_0) \\ 0 & (\text{otherwise}) \end{cases} \quad (3.11)$$

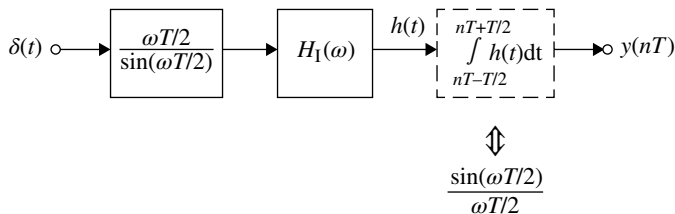
We can see that this is a special case of the transfer function in Equation 3.8, where the roll-off factor  $\alpha=1$ . Thus, the raised cosine roll-off filter with a roll-off factor of  $\alpha=1$  satisfies Nyquist's first and second criteria.

Figure 3.7 shows eye diagrams of the two transfer functions explained earlier given by Equations 3.10 and 3.11. As shown later, a Nyquist's second criterion (Nyquist-II) filter is used for a duobinary transmission system.

### 3.1.3 Nyquist's Third Criterion

This criterion assures that integral of an impulse response over interfering symbol periods is zero and is a constant over the considered time period. It is expressed as following:

$$y(nT) = \int_{nT-T/2}^{nT+T/2} h(t) dt = \begin{cases} \text{constant} & (n = 0) \\ 0 & (n \neq 0) \end{cases} \quad (3.12)$$



**FIGURE 3.8** Nyquist's third criterion filter.

The transfer function is given as:

$$H_{\text{III}}(\omega) = \frac{(\omega T/2)}{(\sin \omega T/2)} H_I(\omega) \quad (3.13)$$

where  $H_I(\omega)$  meets Nyquist's first criterion.

By referring to Figure 3.8, let us examine whether the transfer function in Equation 3.13 meets Nyquist's third criterion. We have shown in Section 2.3.4 that the integral of an input signal for a time range over  $T$  is equivalent to the sampled output of the filter that has the transfer function  $H(\omega) = T \sin(\omega T/2)/(\omega T/2)$ . Thus, the first term on the right side of Equation 3.13 is canceled with  $H(\omega)$ , as far as the sampled signal  $y(nT)$  is concerned. The transfer function  $H_I(\omega)$  ensures  $y(nT) = 0$  ( $n \neq 0$ ). Thus, Equation 3.12 is satisfied.

A Nyquist's third criterion filter is used in a digital frequency modulation (FM) system to generate a modulated signal that takes fixed phases (Section 5.3.2).

### 3.1.4 Other Pulse-Shaping Methods

For a rectangular pulse, the ratio of the pulse time-width to the pulse repetition duration is called the pulse duty ratio. A pulse with the pulse duty ratio of 100% is called an NRZ (nonreturn to zero) signal and other pulses are called RZ (return to zero) signals. An NRZ signal has a rectangular pulse shape such that

$$h(t) = \begin{cases} 1 & \left(-\frac{T}{2} \leq t \leq \frac{T}{2}\right) \\ 0 & \text{(otherwise)} \end{cases}$$

and has a transfer function

$$H(\omega) = T \frac{\sin(\omega T/2)}{\omega T/2}$$

Another frequently used pulse shape has the form of a Gaussian function:

$$h(t) = \frac{1}{\sqrt{2\pi}t_0} e^{-t^2/(2t_0^2)}$$

where  $t_0$  is a constant that represents the width of the pulse. The transfer function also has the Gaussian shape

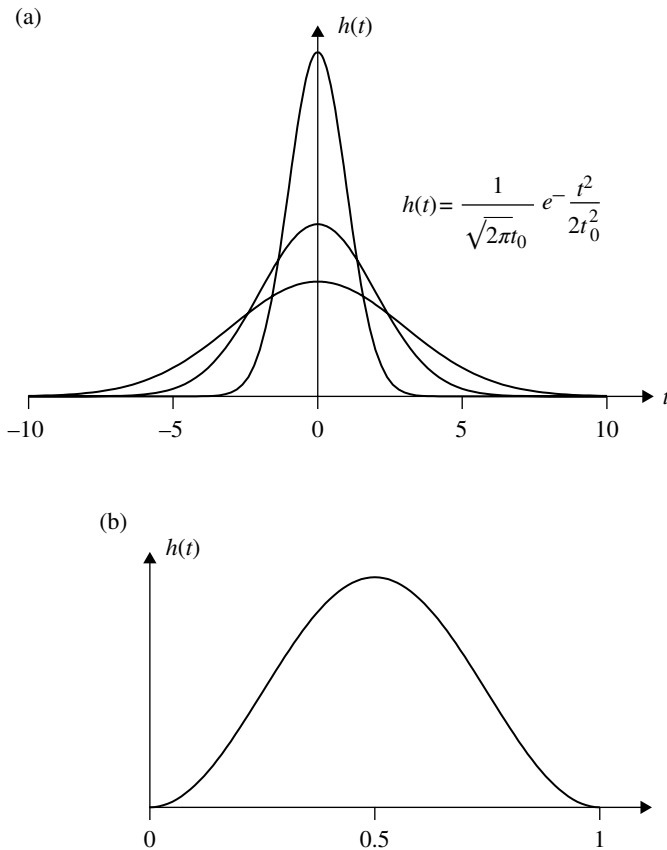
$$H(\omega) = e^{-\frac{\omega^2}{2B_0^2}}$$

where  $B_0 = 1/t_0$  is the bandwidth of the pulse-shaping filter.

Another pulse shaping has a raised cosine waveform and is given as follows:

$$h(t) = \begin{cases} \frac{1}{2LT} \left[ 1 - \cos\left(\frac{2t\pi}{LT}\right) \right] & (0 \leq t \leq LT) \\ 0 & (\text{otherwise}) \end{cases}$$

where  $L$  is an integer. The above two waveforms are shown in Figure 3.9.



**FIGURE 3.9** Pulse waveforms of (a) a Gaussian and (b) the raised cosine impulse response.

## 3.2 LINE CODING

Line coding includes encoding and pulse shaping. Pulse shaping has been described in Section 3.1.

### 3.2.1 Unipolar (On–Off) Code and Polar Codes

A unipolar code assigns a pulse  $p(t)$  to signal “1” and no pulse to “0.” However, a polar code assigns  $p(t)$  to signal “1” and  $-p(t)$  to “0.” Figure 3.10a and b illustrates a unipolar and polar code, respectively, with a pulse duty ratio of 50%. An NRZ unipolar code consists of an NRZ polar code and a dc signal, whose level is one-half of the amplitude of a polar code.

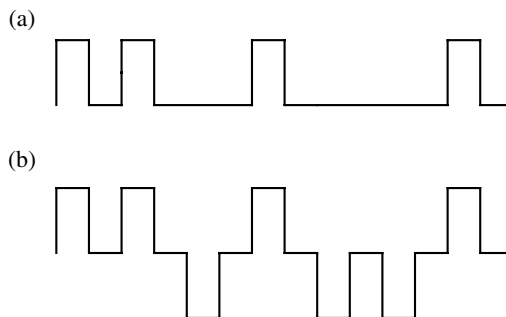
### 3.2.2 Multilevel Codes

A sequence of  $N$  bits signal can be represented by  $2^N$  different states (symbols). A line code with  $2^N$  levels ( $N > 1$ ) is called a *multilevel code*. The symbol frequency  $f_s$  will be  $1/N$  times the bit frequency  $f_b$ . Thus, with the use of  $2^N$ -level codes, the bandwidth of the baseband signal is reduced by a factor of  $1/N$  when compared to the two-level coded system. For example, using 4, 8, and 16 level codes, the bandwidth decreases to  $1/2$ ,  $1/3$ , or  $1/4$  of that of the two-level code.

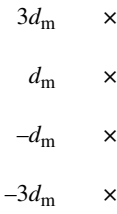
When the average transmit power is kept the same, the minimum distance between the code levels decreases as the number of levels increases. Suppose that the levels are symmetrically spaced with respect to zero voltage (see Fig. 3.11) and that the levels are equally likely to occur, then the minimum distance  $2d_m$  between signal levels will be given as follows:

$$d_m \propto \frac{1}{\sqrt{\frac{2}{2^N} \sum_{n=1}^{2^{N-1}} (2n-1)^2}}$$

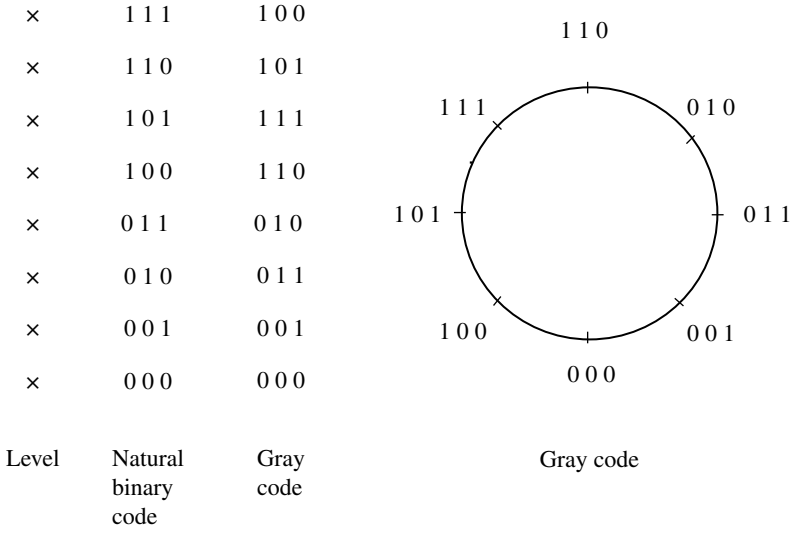
where a  $2^N$ -level code is assumed.



**FIGURE 3.10** (a) Unipolar code and (b) polar code (duty ratio is 50%).



**FIGURE 3.11** Four-level code.



**FIGURE 3.12** Natural binary code and Gray code (eight level).

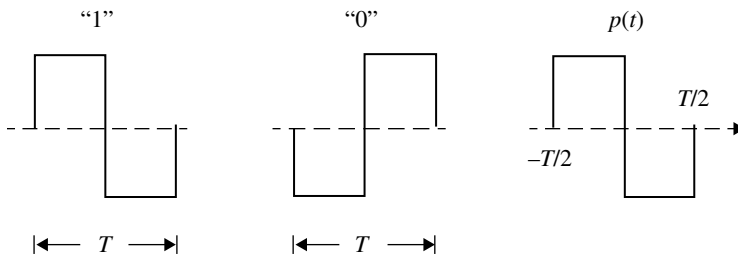
**3.2.3 The Gray Codes**

The one-to-one mapping of  $N$ -bit sequences to  $2^N$  levels is arbitrary. The most straightforward mapping is to use a natural binary code. However, the mapping where the bits assigned to the two neighboring levels differ only by one bit is called *Gray coding*. This property is advantageous in a multilevel transmission system where decision errors occur mostly between two neighboring levels; a symbol error results in a one-bit error in this case, while it may cause a two-bit error or more when a natural binary code is used. Figure 3.12 shows the Gray code and the natural code for an eight-level system. The property of Gray codes that only one bit differs from the neighboring levels is still maintained even when the levels are arranged on a circle.

Consider a  $2^N$ -level ( $N \geq 2$ ) system and express the bit patterns as  $(a_{n-1}, a_{n-2}, a_{n-3}, \dots, a_0)$  for the natural binary code and  $(b_{n-1}, b_{n-2}, b_{n-3}, \dots, b_0)$  for the Gray code. Gray code is generated such that

$$\begin{aligned} b_{n-1} &= a_{n-1} \\ b_i &= a_{i+1} \oplus a_i \quad (i = n-2, n-3, n-4, \dots, 0) \end{aligned} \tag{3.14}$$





**FIGURE 3.13** Manchester (split-phase) code.

where  $\oplus$  denotes the modulo 2 addition, that is,  $0 \oplus 0 = 0$ ,  $0 \oplus 1 = 1$ ,  $1 \oplus 0 = 1$ , and  $1 \oplus 1 = 0$ .

### 3.2.4 Manchester (Split-Phase) Code

The waveforms of this code are shown in Figure 3.13. The spectrum is given as follows:

$$P(\omega) = \int_{-\infty}^{\infty} p(t) e^{-j\omega t} dt = jT \frac{\sin^2(\omega T/4)}{\omega T/4} \quad (3.15)$$

In a symbol duration, the plus and minus areas of each waveform are balanced and, therefore, the spectrum of the signal is null at zero frequency (dc). This property of Manchester codes is advantageous for application to systems where the dc frequency is blocked. Actually, Manchester codes are used in analog (voice) FM systems, in order to transmit the slow-speed data signal in the analog voice channel, which is dc-blocked (i.e., ac-coupled): the voice signal contains no dc component. The ac-coupled circuits are easier to implement because they are free from dc-offset effects.

The other advantage of the Manchester codes is ease of the clock signal extraction (Section 3.4.1) due to the property that polarity of the signal changes in every symbol.

A Manchester code can be seen as binary phase-shift keying where the carrier signal takes a rectangular waveform with a frequency of the bit frequency and is synchronized to the bit clock signal. This property of Manchester codes ensures the use of delay detection, coherent detection, as well as matched filter detection.

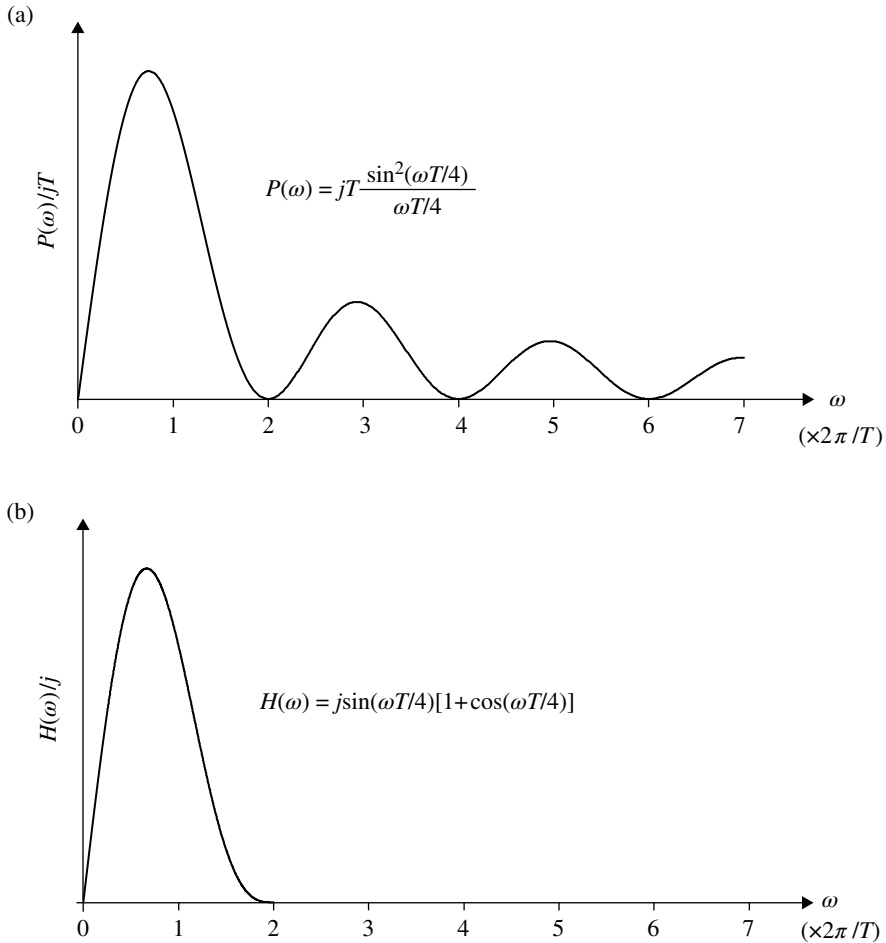
We can band-limit Manchester codes without losing the properties that the spectrum has no dc component and that no intersymbol interference occurs. The idea is that two filters are combined that have impulse responses  $h_1(t)$  and  $h_2(t)$ , respectively,

$$h_1(t) = \delta(t + T/4) - \delta(t - T/4)$$

and  $h_2(t)$  meets the Nyquist's first criterion for the symbol duration of  $T/2$ . Thus, the transfer function of the combined filter is as follows:

$$H(\omega) = H_1(\omega) H_2(\omega) \quad (3.16)$$

where  $h_1(t) \leftrightarrow H_1(\omega) = j2 \sin(\omega T/4)$  and  $H_2(\omega)$  has, for example, the raised cosine roll-off characteristic for the symbol duration of  $T/2$ . The spectrum of a band-limited Manchester code, as given earlier, has a maximum frequency of  $(1 + \alpha)/T$ , where  $\alpha$  is the roll-off factor. Consequently, the bandwidth of Manchester codes is two times the bandwidth of the polar codes. Figure 3.14 shows the transfer functions given by Equations 3.15 and 3.16.



**FIGURE 3.14** Spectra of the Manchester code: (a) without band limitation and (b) with band limitation.

### 3.2.5 Synchronized Frequency Shift Keying Code

As in the case of the Manchester codes, the baseband signal of synchronized frequency-shift keying codes can be presented as a frequency-modulated signal where the carrier signal has a frequency that is comparable with the bit rate frequency and is synchronized to the clock signal. An example of the codes is shown in Figure 3.15. These codes are produced using a rectangular carrier signal with frequency of  $5/4T$  and FM index of 0.5. If we assume a sinusoidal waveform carrier instead of rectangular waveform, the code becomes the minimum shift keying, which is used for data transmission in analog voice FM systems. The modulated signal is matched to systems that have dc-blocked baseband channels.

The mark and the space frequencies that correspond to the digital signal “1” and “0,” respectively, should be chosen so that carrier signal is synchronized with the clock signal. They are given as follows:

$$f_M = N_1 f_B, f_S = N_2 f_B \quad \left( N_1, N_2 = \frac{N}{2}, N = 1, 2, 3, \dots \right)$$

In the case of Figure 3.15,  $N_1 = 1$  and  $N_2 = 3/2$ .

### 3.2.6 Correlative Coding

Correlative coding is a coding scheme where the level of the output (coded) signal is intentionally correlated by manipulating the input digital data. The purpose of correlative coding is to shape the spectrum of the coded signal and to get a narrower or a dc-null spectrum.

The pulse shaping, such as the raised cosine roll-off Nyquist filtering, can be performed independently of the correlative coding. Actually, Nyquist filtering with a roll-off factor of zero is extensively used. Correlative codings fall into two categories. One category is where a nonlinear operation is applied to the input digital data. A number of coding methods of this category are used in the PCM (pulse code modulation) multiplexed transmission system. The other category is where the operation applied to the input signal is linear. Linear coding schemes are discussed in the following.

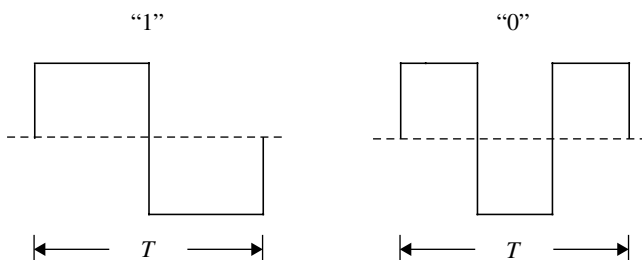


FIGURE 3.15 Synchronized frequency-shift keying code.

**3.2.6.1 Partial Response Codes** This category of codes is generated by passing a polar code into a linear filter with the following impulse response:

$$h(t) = \sum_{n=0}^{N-1} a_n \delta(t - nT)$$

where  $a_n$  is integer constant and  $T$  is the bit duration. The partial response code where  $a_n$  takes real numbers is called the *generalized partial response system*. If the digitally encoded signal is sampled with frequency  $1/T$ , and we denote the operation for a time-delay of  $nT$  by  $z^{-n}$  ( $z$ -transform, Section 2.4.3), then the partial response filter has a  $z$ -transform

$$H(z) = \sum_{n=0}^{N-1} a_n z^{-n}$$

and the output signal spectrum is

$$H(\omega) = \sum_{n=0}^{N-1} a_n e^{-j\omega nT}$$

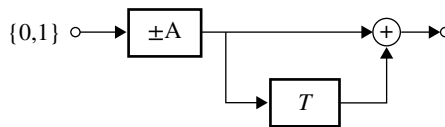
An input pulse produces an output pulse that lasts over a duration of  $N$  symbol periods. The memory length is  $NT$ . In practice, waveform spreads further owing to pulse shaping as in the case of the Nyquist roll-off filtering.

**3.2.6.2 Duobinary (Class I Partial Response) Code** The case of  $N=2$  and  $a_0=a_1=1$  is called duobinary or class I partial response code. The coder is shown in Figure 3.16. It has three levels:  $2A$ ,  $0$ , and  $-2A$ . The transfer function is as follows:

$$H(\omega) = 2 \cos\left(\frac{\omega T}{2}\right) e^{-\frac{j\omega T}{2}} \quad (|\omega| \leq \infty) \quad (3.17)$$

If we use the Nyquist-I filter with a roll-off factor of zero for pulse shaping, then the transmit signal spectrum becomes as follows:

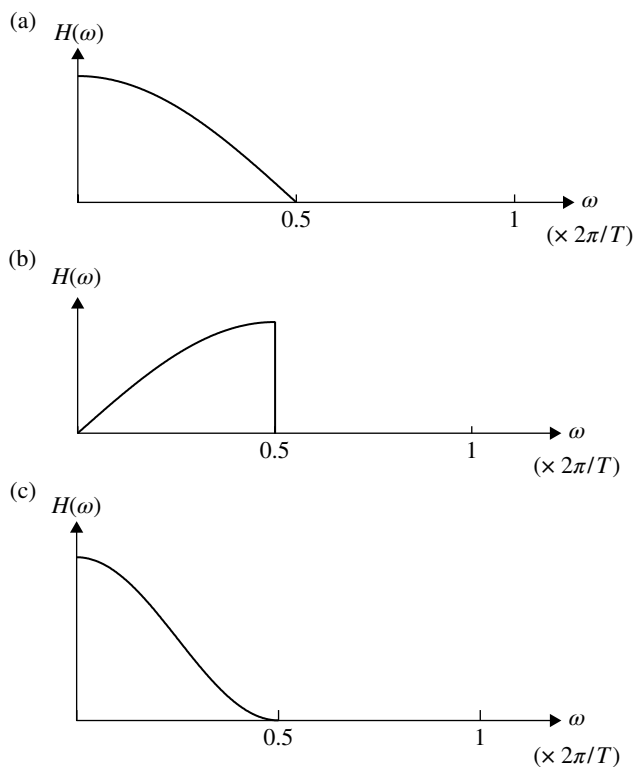
$$G(\omega) = \begin{cases} 2 \cos\left(\frac{\omega T}{2}\right) e^{-\frac{j\omega T}{2}} & \left(0 < |\omega| < \frac{\pi}{T}\right) \\ 0 & \text{(otherwise)} \end{cases}$$



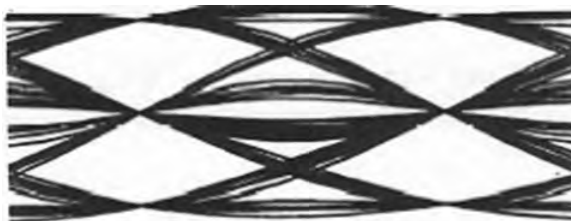
**FIGURE 3.16** Duobinary coding.

The spectrum  $G(\omega)$  is drawn in Figure 3.17a. The eye pattern is shown in Figure 3.18.

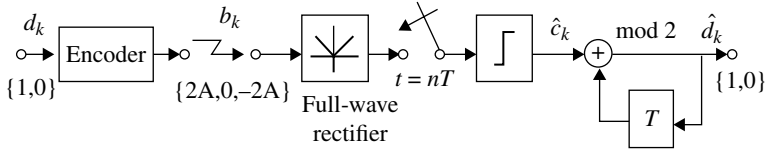
The transmission system with the duobinary code is shown in Figure 3.19. The received signal is full-wave rectified to produce the two-level signal of  $2A$  and  $0$ . Level  $2A$  corresponds to a data “0” and level  $0$  to “1.” The signal determined by the decision-making process is applied to a feedback circuit, where modulo 2 operation is assumed.



**FIGURE 3.17** Spectra of (a) duobinary code, (b) bipolar code, and (c) class II partial response code.



**FIGURE 3.18** Eye diagram of the duobinary code.



**FIGURE 3.19** Duobinary transmission system.

$d_k$	0	1	1	1	0	0	1	0	1
$b_k$		0	2A	2A	0	-2A	0	0	0
$\widehat{c}_k$		1	0	0	1	0	1	1	1
$\widehat{d}_k$	0	1	1	1	0	0	1	0	1
$\widehat{c}'_k$		1	0	0	0	0	1	1	1
$\widehat{d}'_k$		1	1	1	1	1	0	1	0

**FIGURE 3.20** An example of the duobinary transmission process.

Figure 3.20 illustrates the transmission process using an assumed data sequence. Without transmission errors, the output signal  $\widehat{d}_k$  is equal to the transmitted signal  $d_k$ . If an error occurs in the decision process,  $\widehat{c}_k$ , then the output signal  $\widehat{d}_k$  is inverted after this error and the inversion continues until the next error occurs.

Precoding is introduced to prevent the above error propagation in the partial response system (Fig. 3.21). For the duobinary system, the precoding is expressed as follows:

$$d_k = d'_k \oplus d'_{k-1} \pmod{2} \quad \text{or} \quad d'_k = d_k \oplus d'_{k-1} \pmod{2}$$

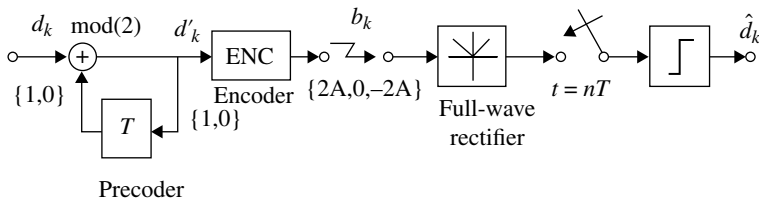
The  $z$ -transform expression of the precoder is as follows:

$$H_p(z) = \frac{1}{1+z^{-1}} \pmod{2}$$

However, the duobinary coding is expressed as follows:

$$H_c(z) = 1 + z^{-1}$$

where arithmetic addition instead of mod 2 addition is assumed.



**FIGURE 3.21** Duobinary transmission system with precoding.

Since levels  $2A$  and  $-2A$  correspond to “0” and the level  $0$  to “1,” the decision process can be seen as performing the mod 2 operation on the duobinary coded signal, that is,

$$H_c(z) = 1 + z^{-1} \pmod{2}$$

Thus, the duobinary transmission system with precoding is interpreted as performing the logical manipulation of

$$H_p(z) \pmod{2} \cdot H_c(z) \pmod{2} = 1$$

on the transmitted signal  $d_k$ .

Comparing Figures 3.19 and 3.21, the receiver feedback circuit is moved to the transmitter as the precoding circuit.

For the  $L$ -level input signal, the precoder is expressed as follows:

$$H_p(z) = \frac{1}{H_c(z)} \pmod{L}$$

where  $H_c(z)$  is given corresponding to the used partial response coding. The precoder and the partial response coder are realized as shown in Figure 3.22. We can see  $H_p(z)H_c(z) = 1$  holds regardless of the type of the input signal and of the modulo operation.

Another method for receiving the duobinary coded signal is shown in Figure 3.23. This detection circuit can be seen as a decision feedback equalizer (Section 7.3.3), where the intersymbol interference caused intentionally by partial response coding is equalized.

**3.2.6.3 Bipolar Code** Bipolar code is also called as *AMI (alternative-mark-inversion) code*. The code assigns pulses with levels  $A$  and  $-A$  to signal “1” and no pulse to signal “0.” The level  $\pm A$  alternates in every “1” as illustrated in Figure 3.24. Thus, the coded waveform is balanced in polarity and has no dc component. The spectrum is shown in Figure 3.17b. The balance is a result of correlating the two successive “1s”. The rule that the levels take alternating polarity can be used for detection of errors by monitoring the violation of the rule.

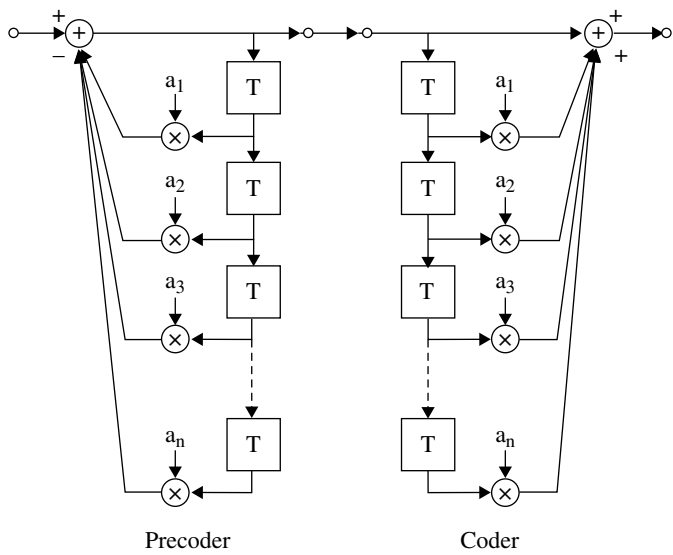


FIGURE 3.22 Precoder and coder for partial response system.

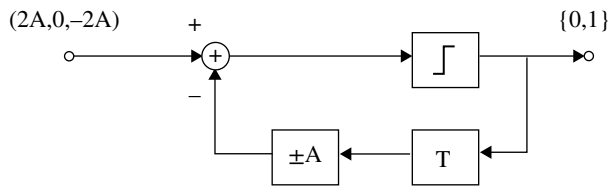


FIGURE 3.23 Decision feedback detection of duobinary coded signals.

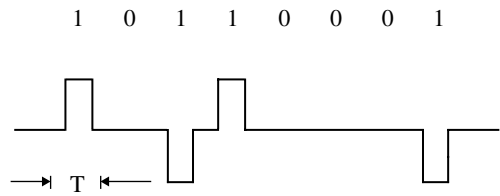


FIGURE 3.24 Bipolar (AMI) coding.

The encoding process has infinite length memory; signal “1” affects the entire output after the signal. A bipolar code can be detected by full-wave rectifying followed by two-level decision. A bipolar code can be generated using a partial response coder

$$H_c(z) = 1 - z^{-1}$$

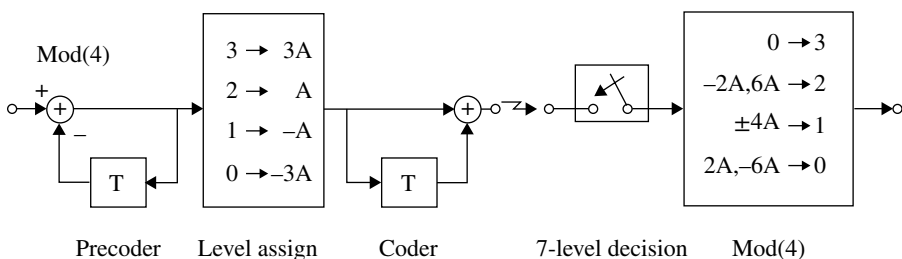


and the precoder  $H_p(z) = 1/H_c(z) \pmod{2}$ . The spectrum of the coded signal then becomes as follows:

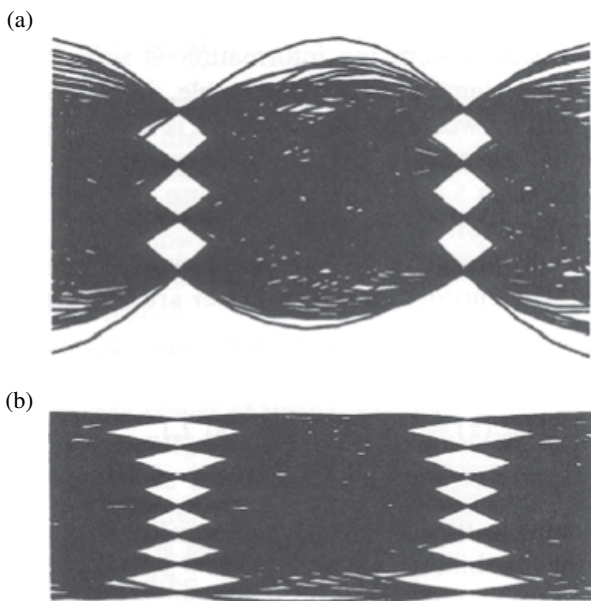
$$H(\omega) = j2 \sin\left(\frac{\omega T}{2}\right) e^{-j\frac{\omega T}{2}} \quad (-\infty < \omega < \infty) \quad (3.18)$$

If we assume the Nyquist-I filter with a roll-off factor of zero, then the spectrum is limited over the frequency range,  $|\omega| \leq \pi/T$  (see Fig. 3.17b).

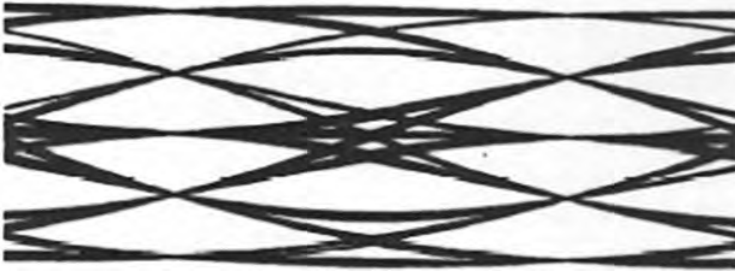
**3.2.6.4 Duoquaternary Code** This is a four-level signal input partial response code with  $H_c(z) = 1 + z^{-1}$ . A duoquaternary transmission system is shown in Figure 3.25. The transmitted signal has seven levels as shown in Figure 3.26b. The seven-level decision is made at the receiver followed by the modulo 4 operation.



**FIGURE 3.25** Duoquaternary transmission system.



**FIGURE 3.26** Eye diagrams of the duoquaternary coded signal. (a) Input four-level signal and (b) duoquaternary signal.



**FIGURE 3.27** Eye diagram of the class II partial response coded signal.

The spectrum of the duoquaternary encoded signal has one-half of the bandwidth of the duobinary signal provided that the bit rate is the same.

**3.2.6.5 Class II Partial Response Code** This code is characterized with  $H_c(z) = 1 + 2z^{-1} + z^{-2}$ . The coded signal takes five levels when a two-level signal input is applied as shown in Figure 3.27. The spectrum is as follows:

$$H(\omega) = 2(3 + \cos \omega T) e^{-j\omega T} \quad (3.19)$$

If we assume the Nyquist-I filter with the roll-off factor to be zero, then the spectrum covers the frequency range  $|\omega| < \pi/T$  (see Fig. 3.17c).

### 3.2.7 Differential Encoding

In a digital transmission system, information is sometimes carried by relative changes of the signal states, for example, in differential detection, the relative phase shift between successive signals is determined. In coherent detection, differential encoding prevents error propagation caused by a carrier phase slip (Section 5.5.1). For this purpose, differential encoding is employed at the transmitter.

The differential encoding circuit and relevant decoding circuit are shown in Figure 3.28. The encoder and the decoder are characterized with the  $z$ -transform,

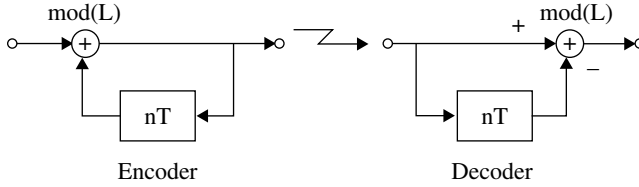
$$H_{\text{enc}}(z) = \frac{1}{1 - z^{-n}} \pmod{L}$$

$$H_{\text{dec}}(z) = 1 - z^{-n} \pmod{L}$$

respectively.

Typically  $n=1$ , and in this case, the differential encoder is the same as the precoder for the duobinary coding. For  $n=2$ , the two-symbol delay differential encoding ( $n=2$ ) is equivalent to repeating the one-symbol delay differential encoding ( $n=1$ ) twice. This can be seen from the fact that  $1 - z^{-2} = (1 - z^{-1})^2 \pmod{2}$ .

It can be seen from Figure 3.28 that a single error in the channel causes two successive errors in the decoding process.



**FIGURE 3.28** The differential encoder and decoder.

### 3.3 SIGNAL DETECTION

The bit error rate is a measure of the quality of the receiver. The error occurs at the decision process due to the noise and interference. Decision-making on the digital signal at the receiver is the crucial difference between digital communication and analog communication. This section describes some fundamental techniques in digital signal detection.

#### 3.3.1 $C/N$ , $S/N$ , and $E_b/N_0$

In order to compare the performances of different modulation and detection methods, a statement of the general condition is desired, which is independent of the actual received power, the bit rate frequency, and the noise figure of the receiver. The energy per bit to noise power ratio,  $E_b/N_0$ , meets this purpose. In the following,  $E_b/N_0$  is expressed by the carrier to noise power ratio  $C/N$  for a modulated signal and by the signal to noise power ratio  $S/N$  for a baseband signal.

For the modulated signal, the energy per bit is defined as follows:

$$E_b = CT_b \quad (3.20)$$

where  $C$  is the average power of the modulated signal measured at the input of the demodulator and  $T_b$  is the bit duration. (The non-modulated carrier power is the same as the modulated signal power with constant envelope modulation and it is different in the non-constant envelope modulation.) When the signal power is measured at the output of the band-pass filter at the receiver, the measured value must take into consideration the signal spectrum and the band-pass filter transfer characteristics. The noise power  $N$  measured at the output of the band-pass filter is related to the noise power density  $N_0$  as

$$N = W_{eq} N_0$$

where  $W_{eq}$  is the equivalent noise bandwidth of the band-pass filter (Section 2.2.1). Thus, we have the following:

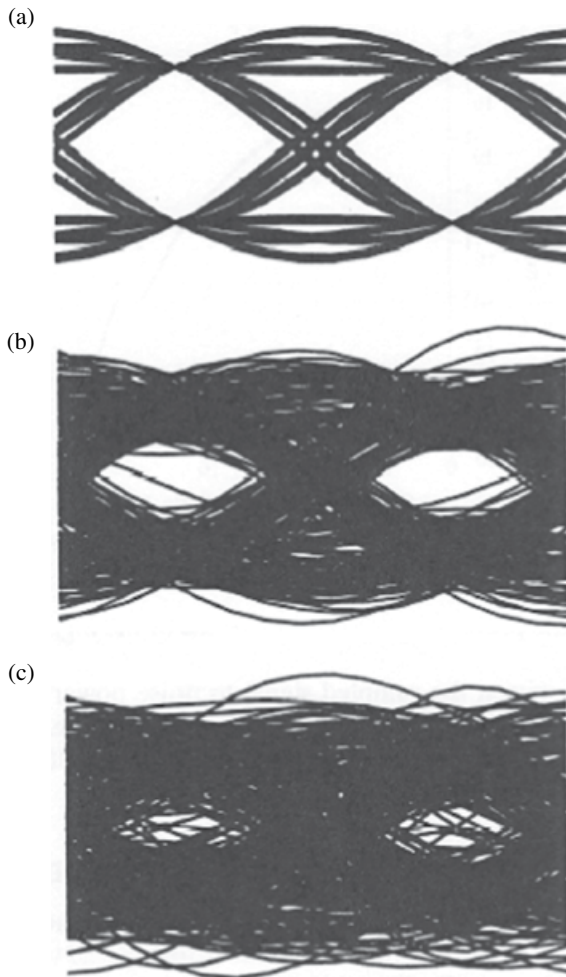
$$\frac{E_b}{N_0} = \frac{CT_b}{N_0} = \frac{W_{eq} C}{f_b N} \quad (3.21)$$

where  $f_b = 1/T_b$  is the bit rate frequency.

For a baseband signal, the above argument can be applied by considering a low-pass filter instead of the band-pass filter. We use a parameter  $C/N_0 = f_b E_b / N_0$  [Hz] when we express the difference in transmit power for different transmission speed. When a matched filter is assumed, we have  $S/N = 2E_b / N_0$  at the sampling instants (Section 3.3.5).

### 3.3.2 Bit Error Rate

A decision error occurs at the sampling instant when noise perturbs the received signal level over the threshold, as shown in Figure 3.29c. We analyze bit error rates in the following.



**FIGURE 3.29** Eye diagram of the two-level polar signal with additive Gaussian noise: (a)  $E_b/N_0 = \infty$ , (b)  $E_b/N_0 = 8 \text{ dB}$ , and (c)  $E_b/N_0 = 4 \text{ dB}$ .

Let us assume a data sequence  $\{d_k\}$  ( $d_k = 0, 1$ ) and signals are transmitted into a channel as follows:

$$x_k = 2d_k - 1$$

$$= \begin{cases} 1 & (d_k = 1) \\ -1 & (d_k = 0) \end{cases}$$

White Gaussian noise with zero mean is added at the receiver. We denote the output of a noise band-limit filter at sampling instant  $k$  as follows:

$$y_k = Ax_k + z_k$$

where  $A$  is a constant and  $z_k$  denotes the Gaussian noise. The probability density function is given by

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2\sigma^2}} \quad (3.22)$$

where  $\sigma^2 = \langle z^2 \rangle$  is the average noise power. The signal to noise power ratio  $S/N$  at a sampling instant is given as follows:

$$\frac{S}{N} = \frac{A^2}{\sigma^2}$$

When the noise band-limit filter is a matched filter (Section 3.3.5), we have:

$$\frac{E_b}{N_0} = \frac{1}{2} \frac{S}{N} \equiv \gamma$$

where  $E_b$  is the energy per bit and  $N_0$  is the noise power density at the receiver input.

Received signal  $y$  (hereafter suffix  $k$  is omitted for simplicity sake) is distributed with a probability density as follows:

$$p(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-Ax)^2}{2\sigma^2}} \quad (3.23)$$

Two-level decision is made with a threshold value  $y_t$  as follows:

$$\hat{x} = \begin{cases} 1 & (y > y_t) \\ -1 & (y < y_t) \end{cases}$$

When transmit data is  $d=1$  ( $x=1$ ), the error rate  $P_1$  is given as follows:

$$P_1(y_t) = \int_{-\infty}^{y_t} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-A)^2}{2\sigma^2}} dy = \int_{(A-y_t)/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y'^2}{2}} dy' = Q\left(\frac{A}{\sigma} - \frac{y_t}{\sigma}\right) \quad (3.24)$$

where  $Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$

Similarly, the error rate for  $d=0$  ( $x=-1$ ) becomes as follows:

$$P_0(y_t) = \int_{y_t}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y+A)^2}{2\sigma^2}} dy = Q\left(\frac{A}{\sigma} + \frac{y_t}{\sigma}\right)$$

We assume the a priori probabilities of data generation as  $P[1]$  ( $d=1$ ),  $P[0]$  ( $d=0$ ). Then, the average error rates become:

$$P_e(y_t) = P[1]P_1(y_t) + P[0]P_0(y_t) \quad (P[1] + P[0] = 1)$$

A threshold value  $y_t$  that gives the minimum average error rate is obtained by differentiating the above equation with respect to  $y_t$  as follows [11]:

$$\begin{aligned} \frac{\partial}{\partial y_t} P_e(y_t) &= P[1] \frac{\partial}{\partial y_t} P_1(y_t) + P[0] \frac{\partial}{\partial y_t} P_0(y_t) \\ &= P[1]p(y_t - A) - P[0]p(y_t + A) = 0 \end{aligned}$$

where  $p(\cdot)$  is given by Equation 3.22. From the above equation, we have:

$$\frac{p(y_t - A)}{p(y_t + A)} = e^{\frac{2Ay_t}{\sigma^2}} = \frac{P[0]}{P[1]}$$

Taking the logarithm of the equation, we get the optimum  $y_t$  value as follows:

$$y_t = \frac{\sigma^2}{2A} \ln \frac{P[0]}{P[1]}$$

This equation is modified as

$$\frac{y_t}{\sigma} = \frac{\sigma}{2A} \ln \frac{P[0]}{P[1]} = \frac{1}{2\sqrt{2\gamma}} \ln \frac{P[0]}{P[1]}$$

Furthermore,

$$\begin{aligned} \frac{A}{\sigma} - \frac{y_t}{\sigma} &= \sqrt{2\gamma} + \frac{1}{2\sqrt{2\gamma}} \ln \frac{P[0]}{P[1]} \\ &= \sqrt{2\gamma} \left(1 + \frac{\lambda}{4\gamma}\right) \end{aligned}$$

where

$$\lambda = -\ln \frac{P[0]}{P[1]} = \ln \frac{P[1]}{P[0]}$$

Thus, the minimum error rate is given as follows:

$$P_e = P[1]Q\left[2\gamma\left(1 + \frac{\lambda}{4\gamma}\right)\right] + P[0]Q\left[2\gamma\left(1 - \frac{\lambda}{4\gamma}\right)\right] \quad (3.25)$$

The above discussion is equivalent to estimate the most probable transmit signal  $x (= \pm 1)$  with a given received signal  $y$ . Therefore, another argument can be developed as follows. The joint probability of  $y$  and  $x$ ,  $P(x, y)$  is given by the a posteriori probability,  $P(y|x)$ , for a given  $x$  and the a priori probability of  $x$ ,  $P(x)$  as follows:

$$P(x, y) = P(y | x) P(x)$$

Therefore, we can get the minimum error rate by calculating the probabilities for  $x=1$  and  $x=-1$  and making a decision for  $x$  that shows higher probability, that is, we check the following inequalities.

$$P(x=1, y) \leq P(x=-1, y)$$

This equation is rewritten as

$$\frac{P(x=1, y)}{P(x=-1, y)} = \frac{P(y | x=1) P(x=1)}{P(y | x=-1) P(x=-1)} \leq 1$$

We define the log-likelihood ratio as

$$\begin{aligned} L(y) &= \log \frac{P(y | x=1) P(x=1)}{P(y | x=-1) P(x=-1)} \\ &= \log \frac{P(y | x=1)}{P(y | x=-1)} + \log \frac{P(x=1)}{P(x=-1)} \end{aligned}$$

From Equation 3.23, we have

$$L_e(y) = \log \frac{P(y | x=1)}{P(y | x=-1)} = \frac{2A}{\sigma^2} y$$

The probability that  $x=1$  ( $d=1$ ) is erroneously decided as  $x=-1$  ( $d=0$ ) is the probability of  $L_e(y) + \lambda = \frac{2A}{\sigma^2} y + \lambda < 0$  and therefore  $y < -\frac{\sigma^2}{2A} \lambda$  for  $p(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-A)^2}{2\sigma^2}}$ .

Then, error rates are given as

$$P_{e1} = \int_{-\infty}^{y_c} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-A)^2}{2\sigma^2}} dx$$

where  $y_t = -(\sigma^2/2A)\lambda$ . The above result is the same as Equation 3.24. Similarly, the error probabilities for  $x=-1$  as  $x=1$  is given as follows:

$$P_{e0} = \int_{y_t}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x+A)^2}{2\sigma^2}} dx$$

As a priori probabilities (information) are usually not known, the threshold is set to zero ( $y_t = 0$ ). Therefore, we have  $P_1(0) = P_0(0) = Q[\sqrt{2\gamma}]$  and error rates never depend on data sequence. The average error rates are given as

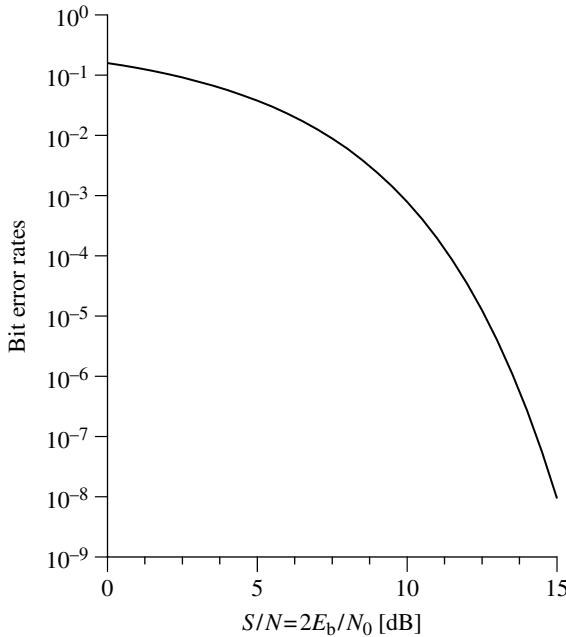
$$P_e = P[1] Q[\sqrt{2\gamma}] + P[0] Q[\sqrt{2\gamma}] = Q[\sqrt{2\gamma}] \quad (3.26)$$

The error rates given in Equation 3.26 are shown in Figure 3.30. For a multilevel transmission, the error rates for the highest or the lowest level signal are one-half of those for the other level signal.

The symbol error rates are given as:

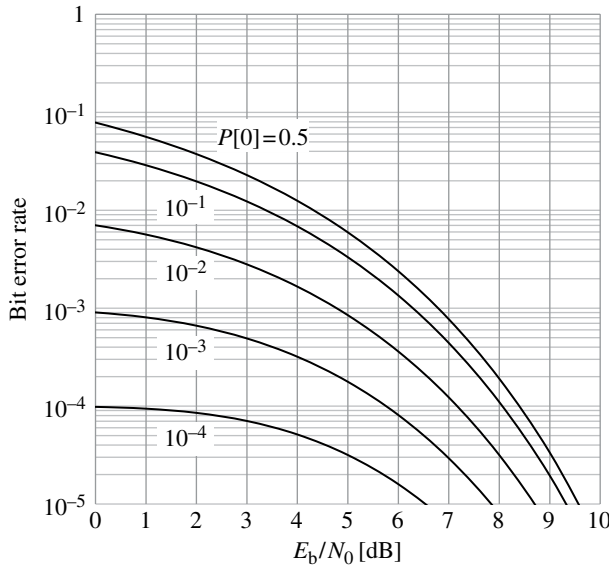
$$P_e = \{2 - P(\pm M)\} Q\left(\frac{d_m}{\sigma}\right) \quad (3.27)$$

where  $P(\pm M)$  is the probability that the highest or the lowest level signal arrives and  $2d_m$  is the distance between the neighboring levels (Fig. 3.11).



**FIGURE 3.30** Bit error rate of the two-level polar signaling system as a function of sampled signal energy to noise power ratio.





**FIGURE 3.31** Bit error rates when a priori probabilities are given.

If the a priori probabilities are known exactly, the average error rates are given with Equation 3.25. Figure 3.31 shows the average error rates with the a priori probability  $P[0]$  as a parameter. The error rates decrease quickly when the a priori probability departs from the equal probabilities ( $P[1]=P[0]=1/2$ ) regardless of  $E_b/N_0$  values. Actually, the exact a priori probabilities are not known, and a threshold value of  $E_b/N_0$  exists, below which the error rate performance improvement cannot be obtained, owing to the Shannon channel capacity (Section 3.8).

The a priori probabilities are estimated using the redundancy of codes such as in turbo codes and low density parity check (LDPC) codes (Sections 7.4.5 and 7.4.6). The error-correcting capability depends on the exactness of the estimation. The estimation is carried out at the receiver. Let us consider the log likelihood ratio  $\lambda = \ln P[1]/P[0]$  as the a priori information. The estimation of  $\lambda$  is made based on the received signal  $y$  (more precisely  $\{y_k\}$ ). We denote the estimated values for  $x=1$  and  $x=-1$  by  $\langle \lambda \rangle|_{x=1}$  and  $\langle \lambda \rangle|_{x=-1}$ , respectively. It is proved that  $\langle \lambda \rangle|_{x=1} = -\langle \lambda \rangle|_{x=-1}$  holds for the LDPC codes [12]. We can prove the same result for the turbo codes.

Therefore, we have the following:

$$\begin{aligned}
 P_e &= P[1]Q\left[\sqrt{2\gamma}\left(1 + \frac{\langle \lambda \rangle|_{x=1}}{4\gamma}\right)\right] + P[0]Q\left[\sqrt{2\gamma}\left(1 - \frac{\langle \lambda \rangle|_{x=-1}}{4\gamma}\right)\right] \\
 &= P[1]Q\left[\sqrt{2\gamma}\left(1 + \frac{\langle \lambda \rangle|_{x=1}}{4\gamma}\right)\right] + P[0]Q\left[\sqrt{2\gamma}\left(1 + \frac{\langle \lambda \rangle|_{x=1}}{4\gamma}\right)\right] \\
 &= Q\left[\sqrt{2\gamma}\left(1 + \frac{\langle \lambda \rangle|_{x=1}}{4\gamma}\right)\right]
 \end{aligned}$$

Thus, the error probabilities for  $1 \rightarrow 0$  and  $0 \rightarrow 1$  are the same for any data sequence. In convention, we assume a data sequence of all ones ( $d_k = 1$  ( $x_k = 1$ ) for any  $k$ ) when we consider a communications system with estimation of the a priori probability. We denote  $\langle \lambda \rangle|_{x=1}$  by  $\lambda$  hereafter. Since  $\lambda$  is estimated on the received signal  $y$ , it becomes a probabilistic variable as well as  $y$ . If we denote the probability density function by  $f(\lambda)$ , the average error rates are given by the following:

$$\langle P_e \rangle = \int_{-\infty}^{\infty} Q \left[ \sqrt{2\gamma} \left( 1 + \frac{\lambda}{4\gamma} \right) \right] f(\lambda) d\lambda$$

### 3.3.3 NRZ Signaling with Integrate-and-Dump Filter Detection

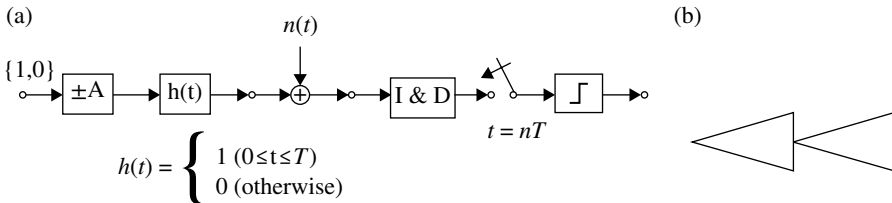
Assume an NRZ signal with amplitude of  $\pm A$  and the bit duration of  $T$ , as shown in Figure 3.32a. The eye diagrams are shown in Figure 3.32b. The energy per bit becomes  $E_b/N_0 = A^2 T$ . The signal with the added white Gaussian noise, whose double-sided noise power density is  $N_0/2$ , is received by the integrate-and-dump (I&D) filter. The signal output of the I&D filter takes levels  $\pm AT$ : the sampled signal power  $S$  is  $(AT)^2$ .

It is shown that at the sampling instant, the signal output of the I&D filter is equivalent to that of the filter with transfer function  $H(\omega)$  given by Equation 2.102. Using this fact and the relation between the average power and the power spectral density given by Equations 2.50 and 2.51, the average noise power  $N$  at the output of the I&D filter becomes

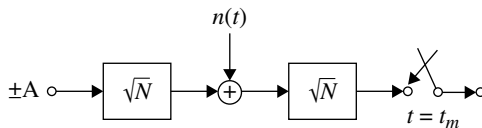
$$N = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{N_0}{2} |H(\omega)|^2 d\omega = \frac{N_0 T}{2} \left( \int_{-\infty}^{\infty} \left( \frac{\sin x}{x} \right)^2 dx = \pi \right)$$

Then, we have

$$\left[ \frac{S}{N} \right]_s = \frac{(AT)^2}{N_0 T / 2} = 2 \frac{E_b}{N_0}$$



**FIGURE 3.32** NRZ signaling with integrate-and-dump filter detection: (a) system model and (b) eye diagram.



**FIGURE 3.33** A two-level Nyquist-I signaling system.  $N$  represents Nyquist's first criterion filter transfer function.

### 3.3.4 Nyquist-I Signaling System

Consider a case where the filter characteristic that meets Nyquist's first criterion is divided equally into transmit and receive filters (Fig. 3.33), for example, the transmit and receive filters have transfer functions  $H_T(\omega)$  and  $H_R(\omega)$ , respectively, which are the square roots of Equation 3.8. We call this as root Nyquist filter. This system is used in all digital cellular systems.

Using Equations 2.58, 3.8, and  $E_b = PT$ , the energy per bit is given as follows:

$$E_b = \frac{A^2}{2\pi} \int_{-\infty}^{\infty} |H_T(\omega)|^2 d\omega = \frac{A^2}{2\pi} \int_{-\infty}^{\infty} |H_1(\omega)| d\omega = \frac{A^2}{T}$$

With the fact that  $H_T(\omega)H_R(\omega) = H_1(\omega)$  and Equation 3.9, the signal at the output of the receive filter takes  $\pm A/T (= \pm Ah(0))$ ; the sampled signal power  $S$  is  $(A/T)^2$ . The noise power at the output of the receive filter becomes

$$N = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{N_0}{2} |H_R(\omega)|^2 d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{N_0}{2} |H_1(\omega)| d\omega = \frac{N_0}{2T}$$

The signal to noise power ratio at the sampling instant becomes  $2A^2T/N_0 = 2E_b/N_0$ . Thus, we have the same results as in the case of NRZ signaling with the I&D filter detection system.

The impulse response  $h(t)$  of the root Nyquist filter is shown by Equation 8.5. It can be shown that  $h(t)$  and its time shift version  $h(t+nT)$  ( $n$ : integer) are orthogonal with each other as shown in the following equation (Section 2.2.4):

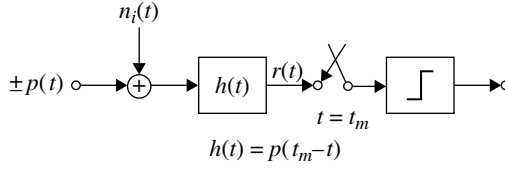
$$\int_{-\infty}^{\infty} h(t)h(t+nT)dt = 0$$

The fact that noises at different sampling instants are independent (Section 2.2.4) is another reason why the root Nyquist filter is usually employed.

### 3.3.5 The Matched Filter

The filter that maximizes the signal to noise power ratio at the sampling instant is known as *the matched filter*. Consider the case that a signal  $p(t)$  with added noise  $n_i(t)$  is applied to a filter (Fig. 3.34). The power spectral density of the input noise  $n_i(t)$  is denoted by  $G(\omega)$ . The signal output from the filter at the time  $t_m$  is given as follows:

$$r(t_m) = p(t) * h(t)|_{t=t_m} = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega)H(\omega)e^{j\omega t_m}d\omega$$

**FIGURE 3.34** Matched filter receiver.

where  $h(t)$  is the filter impulse response, “ $*$ ” denotes the convolution integral, and  $p(t) \leftrightarrow P(\omega), h(t) \leftrightarrow H(\omega)$ .

The average filtered noise power is

$$\langle n^2(t_m) \rangle = \langle |n_i(t) * h(t)|^2_{t=t_m} \rangle$$

If noise is stationary,  $\langle n^2(t_m) \rangle$  never depends on time  $t$ . We assume the ensemble and the time average are equal, and thus we have:

$$\langle n^2(t_m) \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) |H(\omega)|^2 d\omega$$

The signal to noise power ratio at the sampling instant  $t = t_m$  becomes as follows:

$$\frac{S}{N} = \frac{r^2(t_m)}{\langle n^2(t_m) \rangle} = \frac{\left| \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega) H(\omega) e^{j\omega t_m} d\omega \right|^2}{\frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) |H(\omega)|^2 d\omega}$$

To get the optimum transfer function  $H(\omega)$ , which maximizes this value, we use the Schwartz inequality, viz.,

$$\int_{-\infty}^{\infty} |X(\omega)|^2 d\omega \int_{-\infty}^{\infty} |Y(\omega)|^2 d\omega \geq \left| \int_{-\infty}^{\infty} X(\omega) Y(\omega) d\omega \right|^2 \quad (3.28)$$

where the equality holds when  $X(\omega) = kY^*(\omega)$  ( $k$ : constant).

With  $X(\omega) = \sqrt{G(\omega)} H(\omega) e^{j\omega t_m}$  and  $Y(\omega) = P(\omega) / \sqrt{G(\omega)}$ , we get

$$\frac{S}{N} \leq \frac{1}{2\pi} \frac{\int_{-\infty}^{\infty} \frac{|P(\omega)|^2}{G(\omega)} d\omega \int_{-\infty}^{\infty} |\sqrt{G(\omega)} H(\omega) e^{j\omega t_m}|^2 d\omega}{\int_{-\infty}^{\infty} G(\omega) |H(\omega)|^2 d\omega} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|P(\omega)|^2}{G(\omega)} d\omega$$

with

$$H_m(\omega) = k \frac{P^*(\omega)}{G(\omega)} e^{-j\omega t_m}$$

For the simplicity of the argument, assume white noise with  $G(\omega) = N_0/2$ . The maximum  $[S/N]_m$  becomes

$$\left[ \frac{S}{N} \right]_m = \frac{2E_b}{N_0} \quad (3.29)$$

where  $E_b$  denotes the pulse energy given as

$$E_b = \int_{-\infty}^{\infty} p^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |P(\omega)|^2 d\omega$$

We have also

$$H_m(\omega) = P^*(\omega) e^{-j\omega t_m}$$

where  $k = N_0/2$  is assumed.

Taking the inverse Fourier transform of  $H_m(\omega)$ , we have

$$\begin{aligned} h_m(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} H_m(\omega) e^{j\omega t} d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} P^*(\omega) e^{j\omega(t-t_m)} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(-\omega) e^{j\omega(t-t_m)} d\omega \\ &= p(t_m - t) \end{aligned} \quad (3.30)$$

where  $h_m(t) \leftrightarrow H_m(\omega)$ .

Thus, the impulse response of the matched filter is given by flipping the time axis and delaying the pulse waveform by  $t_m$ . The signal output  $y(t)$  of the matched filter becomes as follows:

$$y(t) = \int_{-\infty}^{\infty} r(\tau) h_m(t - \tau) d\tau = \int_{-\infty}^{\infty} r(\tau) p(t_m - t + \tau) d\tau$$

Then, at the sampling instant  $t = t_m$ ,

$$y(t_m) = \int_{-\infty}^{\infty} r(\tau) p(\tau) d\tau \quad (3.31)$$

If there is no noise,

$$y(t_m) = \int_{-\infty}^{\infty} p^2(\tau) d\tau \quad (3.32)$$

For an example of a matched filter response, consider the input signal

$$p(t) = A\delta(t) + B\delta(t - t_b)$$

The system under consideration can be seen in Figure 3.35a. We have

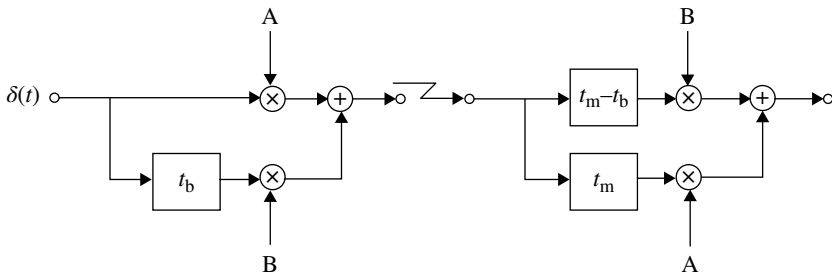
$$y(t) = AB\delta(t - t_m + t_b) + (A^2 + B^2)\delta(t - t_m) + AB\delta(t - t_m - t_b)$$

as in Figure 3.35b. We get the pulse peak at the time instant  $t_m$ , which can be set arbitrarily as long as  $t_m > t_b$ . The pulse peak never depends on the sign of the input signal amplitude and is generated by collecting the spread input signals. This is the principle of a rake receiver in a spread spectrum system (Section 6.4).

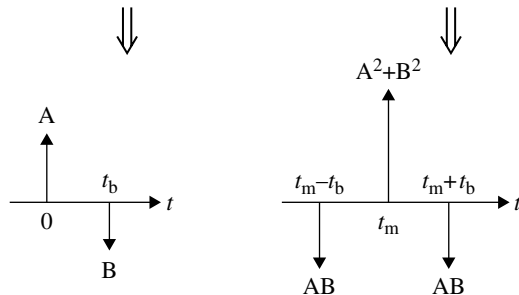
The previously discussed receive filters, that is, I&D filter for NRZ signals and the receive filter for a Nyquist-I transmission system, are both matched filters. When applying a matched filter to a digital transmission system, we must take the intersymbol interference into consideration. Both of the above filters meet the requirements of the matched filter and of intersymbol interference-free transmission. This is not always true.

A matched filter never causes intersymbol interference if the input pulse is limited in a time range less than the symbol duration. In this case, although the output signal

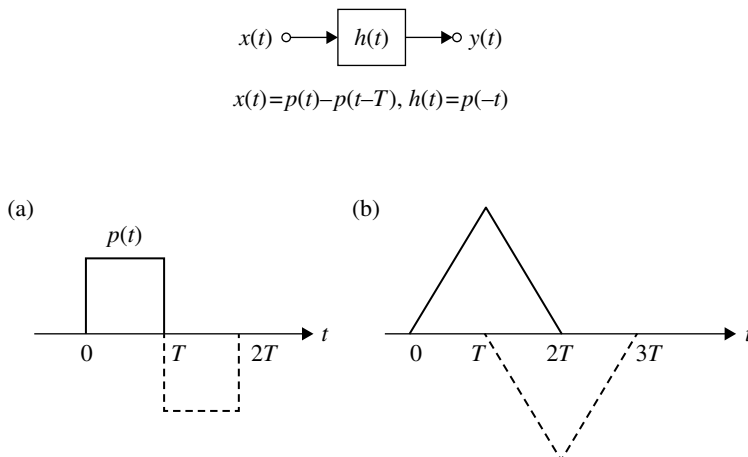
(a) System model



(b) Response



**FIGURE 3.35** An example of a matched filter system.



**FIGURE 3.36** Matched filter response for a pulse waveform limited within a symbol duration. (a) Input and (b) output.

from the matched filter is spread over twice the time range of a pulse duration, it becomes null at the next sampling instant, as illustrated in Figure 3.36. Even though the input pulse is not limited in the time range of the bit duration, the intersymbol interference can be avoided if

$$\begin{aligned}
 y(t_m - nT) &= \int_{-\infty}^{\infty} p(\tau) h(t_m - nT - \tau) d\tau = \int_{-\infty}^{\infty} p(\tau) p(t + nT) d\tau \\
 &= \begin{cases} \int_{-\infty}^{\infty} p^2(\tau) d\tau & (n=0) \\ 0 & (n \neq 0) \end{cases} \quad (3.33)
 \end{aligned}$$

As previously discussed, this is a case where the receive-matched filter with the transmit filter as a whole meets the Nyquist-I criterion. When symbol by symbol instantaneous decisions are assumed, a matched filter receiver without intersymbol interference is optimum for attaining the lowest bit error rate as will be discussed later. It is worthy to note that the pulse shape can be arbitrary as long as the pulse energy is the same (Eq. 3.29).

Equation 3.31 suggests another implementation of the receiver that is known as a *correlation receiver* (Fig. 3.37). In this system, the locally generated signal  $p(t)$  is multiplied to and is synchronized with the incoming signal  $r(t)$ . The output signal at the sampling instant  $t_m$  is the same as that of the matched filter. An NRZ signaling system with an I&D filter detection is a correlation receiver. The matched filter receiver is superior to the correlation receiver in an engineering sense, since it requires no synchronization of the local signal to the incoming signal.

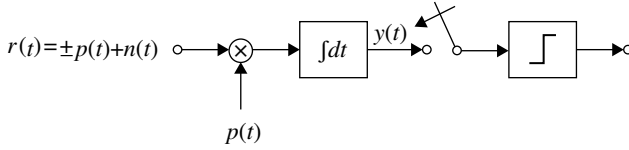


FIGURE 3.37 Correlation receiver.

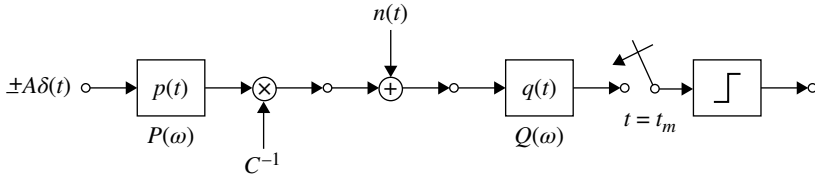


FIGURE 3.38 Joint optimization of transmit and receive filters.

### 3.3.6 Joint Optimization of the Transmit and the Receive Filters

The aforementioned arguments deal with the receive filtering assuming that the transmit filtering is given. This section discusses the optimization of the transmit and receive filters together. The results of the optimization can be different depending on the assumed conditions. We assume that (i) the transmitted signal power is kept constant and (ii) the intersymbol interference is absent at the sampling instant. Under these conditions, the transmit and receive filter combination that maximizes the signal to noise power ratio at the sampling instants will be given. Figure 3.38 shows the system to be optimized. From assumption (ii), the transfer functions  $P(\omega)$  and  $Q(\omega)$  are related as follows:

$$P(\omega)Q(\omega) = H_1(\omega) \quad (3.34)$$

where  $H_1(\omega)$  fulfills the Nyquist-I criterion and  $(h_1(t) \leftrightarrow H_1(\omega))$  takes the peak value at  $t = t_m$ ; it follows that

$$H_1(\omega) = |H_1(\omega)|e^{-j\omega t_m} \quad (3.35)$$

A coefficient  $C^{-1}$  is introduced in order to keep the transmitted signal power to be constant irrespective of the transmit filter characteristics.

The transmitted power  $P_t$  is given as Equation 2.58:

$$P_t = \frac{A^2}{2\pi T} \int_{-\infty}^{\infty} |P(\omega)|^2 d\omega \quad (3.36)$$

where  $A$  is the pulse amplitude and  $T$  is the symbol duration. The coefficient  $C^{-1}$  is chosen such as

$$C^{-1} \equiv \frac{1}{C} = \frac{A}{\sqrt{P_t}} \quad (3.37)$$



At the sampling instant ( $t=t_m$ ), the receive-filter output signal power becomes as follows:

$$S = \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} A P(\omega) C^{-1} Q(\omega) e^{j\omega t_m} d\omega \right|^2$$

With Equations 3.34 and 3.35, this is reduced to the following:

$$S = \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} A C^{-1} |H_1(\omega)| d\omega \right|^2$$

The output noise power becomes

$$N = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) |Q(\omega)|^2 d\omega$$

where  $G(\omega)$  is the noise power spectral density.

With Equations 3.36 and 3.37, the signal to noise power ratio is given as follows:

$$\frac{S}{N} = \frac{\left| \frac{A}{2\pi} \int_{-\infty}^{\infty} |H_1(\omega)| d\omega \right|^2}{\frac{1}{2\pi T} \int_{-\infty}^{\infty} |P(\omega)|^2 d\omega \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) |Q(\omega)|^2 d\omega}$$

$S/N$  is maximized when the denominator takes the minimum value. Using the Schwartz inequality (Eq. 3.28) with  $X(\omega)=P(\omega)$  and  $Y(\omega)=\sqrt{G(\omega)}Q(\omega)e^{j\omega t_m}$ , the minimum value is obtained when

$$P(\omega) = G(\omega)^{1/2} Q^*(\omega) e^{-j\omega t_m} \quad (3.38)$$

Using Equations 3.34, 3.35, and 3.38, we get the following:

$$\begin{aligned} |P(\omega)|^2 &= G(\omega)^{1/2} |H_1(\omega)|, \\ |Q(\omega)|^2 &= G(\omega)^{-1/2} |H_1(\omega)| \end{aligned}$$

The maximum  $S/N$  becomes

$$\left[ \frac{S}{N} \right]_m = \frac{\left| \frac{A}{2\pi} \int_{-\infty}^{\infty} |H_1(\omega)| d\omega \right|^2}{\frac{1}{T} \left| \int_{-\infty}^{\infty} G(\omega)^{1/2} |H_1(\omega)| d\omega \right|^2}$$

Assuming a white noise with  $G(\omega) = N_0/2 (-\infty < \omega < \infty)$ , we have

$$|P(\omega)|, |Q(\omega)| \propto \sqrt{|H_1(\omega)|}$$

and

$$\left[ \frac{S}{N} \right]_m = \frac{2A^2T}{N_0} = \frac{2E_b}{N_0}$$

where  $E_b = A^2T$  is the transmitted energy per bit. These results are the same as those for the previously discussed Nyquist-I signaling system.

### 3.3.7 The Optimum Receiver

The fact that the digital signal has a finite number of states makes possible a decision process with finite number of calculations. Finding the most probable message is possible, even if the received message is corrupted by noise and/or interference.

Let us consider the message expressed as  $\mathbf{m}_i = (m_{i1}, m_{i2}, m_{i3}, \dots, m_{in})$ , where  $m_{in}$  ( $n = 1, 2, \dots, N$ ) is the digital signal to be transmitted. The transmitted signal is given as follows:

$$s_j(t) = \sum_{n=1}^N a_{in} h(t - (n-1)T) \quad (0 < t < T_m, i = 1, 2, \dots, I_m)$$

where  $a_{in}$  takes one of the  $L$  levels corresponding to  $m_{in}$ ,  $h(t)$  is the impulse response of the transmit filter,  $T$  the symbol duration,  $N$  the message length,  $I_m$  the number of different messages, and  $T_m$  the time length of the message. If the message length is  $N$ , then  $I_m = L^N$ . In this case, we can choose the most probable message from these possible candidates by  $L^N$  calculations of the objective function for decision.

Received message signal can be expressed as

$$r(t) = u_i(t) + n(t) \quad (0 < t < T_0 \equiv T_m + \tau_0)$$

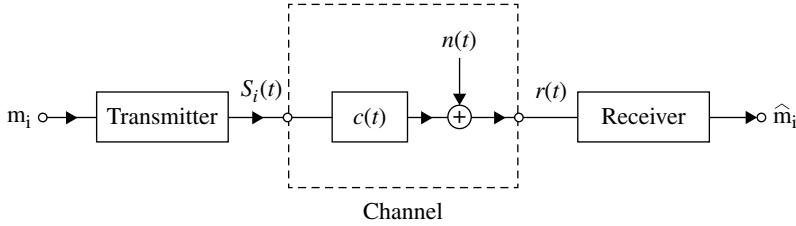
Where

$$u_i(t) = \sum_{n=1}^N a_{in} g(t - (n-1)T_s) \quad (3.39)$$

$g(t)$  denotes the received pulse waveform, which is produced by applying the pulse waveform  $h(t)$  to the channel; that is,  $g(t) = h(t) * c(t)$ , where  $c(t)$  is the impulse response of the channel, assumed to last for  $\tau_0$  time, and  $n(t)$  denotes the noise added in the channel.

Given the received signal  $r(t)$ , our goal is to find the optimum receiver, which yields the lowest error probability for the received candidate messages  $\mathbf{m}_i$ . Figure 3.39 shows the system under consideration.

We proceed our discussion by referring to Refs. [4, 5]. Complex signals are treated in this section. Here we assume that the channel impulse response is known. (Estimating the channel impulse response is another problem, which will be discussed

**FIGURE 3.39** Digital transmission system.

later.) Furthermore, we assume that the probability of the occurrence of each message is known and the noise is Gaussian with zero mean.

When a message  $\mathbf{m}_i$  is sent, the received signal  $r(t)$  is a Gaussian variable with the mean value of  $u_i(t)$ . In order to analyze the stochastic process, it is convenient to represent the received signal with independent variables. The method of treatment is the Karhunen–Loève series expansion [10].

We expand  $n(t)$  in a series in the following form:

$$n(t) = \lim_{N \rightarrow \infty} \sum_{n=1}^N n_n f_n(t) \quad (3.40)$$

where  $n_n$  is the expansion coefficient and  $\{f_n\}$  is a set of orthonormal functions on the interval  $(0, T_0 \equiv T_m + \tau_0)$ , that is,

$$\int_0^{T_0} f_n(t) f_m^*(t) dt = \begin{cases} 1 & (n = m) \\ 0 & (n \neq m) \end{cases} \quad (3.41)$$

Multiplying both sides of the Equation 3.40 by  $f_m^*(t)$ , and using the above relation, we have

$$n_m = \int_0^{T_0} n(t) f_m^*(t) dt \quad (3.42)$$

We impose the condition that the coefficients  $\{n_m\}$  are mutually uncorrelated, that is,

$$\langle n_m n_m^* \rangle = \begin{cases} |\sigma_n|^2 & (m = n) \\ 0 & (m \neq n) \end{cases}$$

Consequently, the coefficients  $n_n$  are independent Gaussian variables with variance  $\sigma_n^2$ . This condition can be filled if the orthonormal functions  $f_n(t)$  satisfy the following integral equation [10]:

$$\int_0^{T_0} R(t, s) f_n(s) ds = |\sigma_n|^2 f_n(t) \quad (3.43)$$

where  $R(t, s)$  is the correlation function defined as

$$R(t, s) = \langle n(t) n^*(s) \rangle = \left\langle \sum_{m=1}^{\infty} n_m f_m(t) \sum_{n=1}^{\infty} n_n^* f_n^*(s) \right\rangle$$

We can confirm that the random variables  $n_n$  are uncorrelated, as follows. Under the condition that Equation 3.43 holds, using Equation 3.42, we have

$$\begin{aligned} \langle n_n n_m^* \rangle &= \left\langle \int_0^{T_0} n(t) f_n^*(t) dt \int_0^{T_0} n^*(s) f_m(s) ds \right\rangle \\ &= \int_0^{T_0} \int_0^{T_0} R(t, s) f_n^*(t) f_m(s) dt ds \\ &= \int_0^{T_0} f_n^*(t) |\sigma_m|^2 f_m(t) dt \\ &= \begin{cases} |\sigma_n|^2 & (n = m) \\ 0 & (n \neq m) \end{cases} \end{aligned}$$

In the language of integral equations, the functions  $f_k(s)$  are eigen (or characteristic) functions and the values  $|\sigma_k|^2$  are eigen (or characteristic) values.

In general, it is difficult to solve Equation 3.43 to get the eigen functions  $f_k(t)$  and the eigen values  $|\sigma_k|^2$ , unless  $n(t)$  is white noise. By assuming white noise, the generality of the argument holds by introducing the concept of a noise-whitening filter.

For the white noise, we have

$$R(t, s) = \frac{N_0}{2} \delta(t - s)$$

where  $\delta(\cdot)$  is the delta function and  $N_0$  is the noise power density. Applying this equation to Equation 3.43, we get the following result:

$$\frac{N_0}{2} f_k(t) = |\sigma_k|^2 f_k(t)$$

This means that any orthonormal function  $f_k(t)$  can be used with

$$|\sigma_k|^2 = \frac{N_0}{2} \equiv \sigma_0^2$$

as its eigen values.

Now, we expand the signals  $u_i(t)$  in terms of orthonormal functions  $f_k(t)$ . The dimension of the signals  $u_i(t)$  is  $N$ . Thus, the signals can be expanded in terms of the orthonormal functions as [4] follows:

$$u_i(t) = \sum_{n=1}^N u_{in} f_n(t) \quad (i = 1, 2, \dots, I_m) \quad (3.44)$$

where the expansion coefficient  $u_{in}$  is given as

$$u_{in} = \int_0^{T_0} u_i(t) f_n^*(t) dt$$

The noise can be expressed as

$$n(t) = n_N(t) + n_0(t),$$

where

$$n_N(t) = \sum_{n=1}^N n_n f_n(t)$$

$$n_0(t) = \sum_{n=N+1}^{\infty} n_n f_n(t)$$

We can see that the noise term  $n_0(t)$  is orthogonal to the signals  $u_i(t)$ , that is:

$$\int_0^{T_0} n_0(t) u_i^*(t) dt = 0$$

Thus, the noise term  $n_0(t)$  is irrelevant to the decision-making at the receiver and can be discarded.

We can express the received signals as follows:

$$r(t) = \sum_{n=1}^N r_n f_n(t) \quad (3.45)$$

where

$$r_n = \int_0^{T_0} r(t) f_n^*(t) dt = u_{in} + n_n \quad (i = 1, 2, \dots, I_m, n = 1, 2, \dots, N)$$

$r_n$  are independent Gaussian variables with mean  $u_{in}$ . Since there is one-to-one correspondence between the received signal  $r(t)$  and the coefficients  $r_n$ , we express the received signal as a vector by  $\mathbf{r} = (r_1, r_2, \dots, r_N)$ .

Given that  $\mathbf{r} = \mathbf{r}'$ , the conditional probability of the correct decision of  $\mathbf{m}_i$  becomes

$$P(C | \mathbf{r} = \mathbf{r}') = P(\mathbf{m}_i | \mathbf{r} = \mathbf{r}')$$

where  $P(C | \mathbf{r} = \mathbf{r}')$  is the conditional probability of making the correct decision given that  $\mathbf{r} = \mathbf{r}'$ .  $P(\mathbf{m}_i | \mathbf{r} = \mathbf{r}')$  is the conditional probability that the message  $\mathbf{m}_i$  was sent, given that  $\mathbf{r} = \mathbf{r}'$  was received. Our task is to maximize  $P(\mathbf{m}_i | \mathbf{r} = \mathbf{r}')$  or to find the message  $\mathbf{m}_i$  for which  $P(\mathbf{m}_i | \mathbf{r} = \mathbf{r}') > P(\mathbf{m}_j | \mathbf{r} = \mathbf{r}')$  for all  $j \neq i$ . The probability  $P(\mathbf{m}_i | \mathbf{r} = \mathbf{r}')$  is called the a posteriori probability of  $\mathbf{m}_i$ . Our receiver is the maximum a posteriori probability detector.

We use Bayes' mixed rule to get the a posteriori probabilities as

$$P(\mathbf{m}_i | \mathbf{r} = \mathbf{r}') = \frac{P(\mathbf{m}_i)P(\mathbf{r} = \mathbf{r}' | \mathbf{m}_i)}{P(\mathbf{r} = \mathbf{r}')} \quad (3.46)$$

$P(\mathbf{r} = \mathbf{r}' | \mathbf{m}_i)$  is the conditional probability that  $\mathbf{r} = \mathbf{r}'$ , given that the message  $\mathbf{m}_i$  was sent. The probability  $P(\mathbf{r} = \mathbf{r}')$  is common to all the decision candidates, so it can be ignored. Our task is to find a candidate message  $\mathbf{m}_i$ , which maximizes the numerator of the right-hand side of Equation 3.46. We denote this term by

$$J_i = P(\hat{\mathbf{m}}_i)P(\mathbf{r} = \mathbf{r}' | \hat{\mathbf{m}}_i) \quad (i = 1, 2, \dots, I_m)$$

Since  $r_n (= r_n - u_{in})$  are Gaussian variables, then

$$P(r_n = r'_n | \hat{\mathbf{m}}_i) = \frac{1}{(2\pi\sigma_0^2)^{1/2}} \exp\left(-\frac{|r'_n - \hat{u}_{in}|^2}{2\sigma_0^2}\right)$$

Since they are independent, then the joint probability function for  $\mathbf{r}$  is as follows:

$$P(\mathbf{r} | \hat{\mathbf{m}}_i) = \frac{1}{\prod_{n=1}^N (2\pi\sigma_0^2)^{\frac{1}{2}}} \exp\left(\sum_{n=1}^N -\frac{|r_n - \hat{u}_{in}|^2}{2\sigma_0^2}\right)$$

where  $\hat{u}_{in}$  corresponds to  $\hat{\mathbf{m}}_i$ .

Since

$$|r_n - \hat{u}_{in}|^2 = |r_n|^2 - 2\text{Re}\{r_n \hat{u}_{in}^*\} + |\hat{u}_{in}|^2$$

then using Equations 3.41, 3.44, and 3.45, we have

$$\begin{aligned} \sum_{n=1}^N |r_n|^2 &= \int_0^{T_0} |r(t)|^2 dt, \\ \sum_{n=1}^N r_n \hat{u}_{in}^* &= \int_0^{T_0} r(t) \hat{u}_i^*(t) dt, \end{aligned}$$

and

$$\sum_{n=1}^N |\hat{u}_{in}|^2 = \int_0^{T_0} |\hat{u}_i(t)|^2 dt$$

Thus, we have

$$\sum_{n=1}^N |r_n - \hat{u}_{in}|^2 = \int_0^{T_0} |r(t) - \hat{u}_i(t)|^2 dt$$

If  $P(\mathbf{m}_i)$  are the same, our task can be reduced to finding a candidate message that minimizes the squared error,  $\int_0^{T_0} |r(t) - \hat{u}(t)|^2 dt$ .

Since the logarithmic function is monotonic, maximizing  $J_i$  is equivalent to maximizing  $\ln(J_i)$ , where  $\ln(\cdot)$  is the natural logarithmic function. Using this fact, we get

$$\ln(J_i) = Q_i + a \quad i = 1, 2, \dots, I_m$$

where

$$\begin{aligned} Q_i &= \ln\{P(\hat{\mathbf{m}}_i)\} - \frac{1}{2\sigma_0^2} \int_0^{T_0} |r(t) - \hat{u}_i(t)|^2 dt, \\ a &= -\frac{N}{2} \ln(2\pi\sigma_0^2) \end{aligned} \quad (3.47)$$

The constant term is common to all the candidate messages and can be ignored in decision-making. Multiplying both sides of Equation 3.47 by  $2\sigma_0^2 (= N_0)$ , we have

$$U_i = N_0 \ln\{P(\hat{\mathbf{m}}_i)\} - \int_0^{T_0} |r(t) - \hat{u}_i(t)|^2 dt \quad (i = 1, 2, \dots, I_m) \quad (3.48)$$

We can rewrite Equation 3.48 as

$$U_i = 2 \operatorname{Re} \left[ \int_0^{T_0} r(t) \hat{u}_i^*(t) dt \right] - e_i - c + N_0 \ln\{P(\hat{\mathbf{m}}_i)\}$$

where

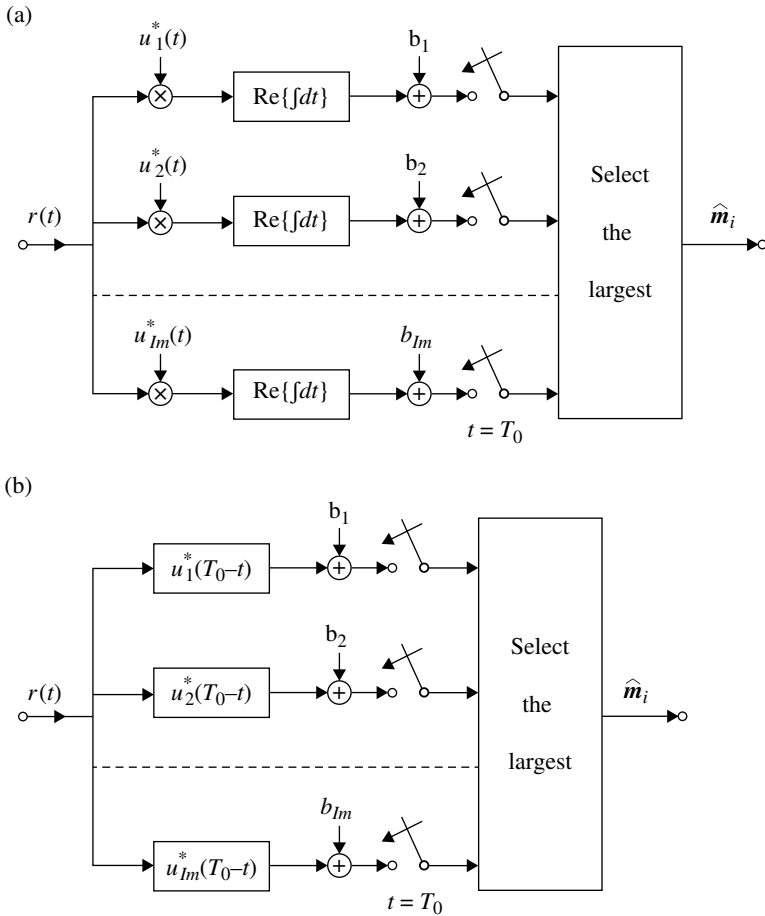
$$e_i = \int_0^{T_0} |\hat{u}_i(t)|^2 dt, \quad c = \int_0^{T_0} |r(t)|^2 dt$$

The term  $e_i$  represents the energy of the candidate signal  $u_i(t)$ ; the term  $c$  is common to all the candidates and can be discarded.

Thus, the optimum receiver is configured with a bank of correlators as shown in Figure 3.40a, where the term  $b_i$  equals  $\{N_0 \ln\{P(\hat{\mathbf{m}}_i)\} - e_i\}/2$ . Since the sampled output of the matched filter is equivalent to the output of the correlator (Section 3.3.5), we have another configuration of the optimum receiver as shown in Figure 3.39b.

Let us recall that  $u_i(t)$  is expanded in terms of the orthonormal functions as in Equation 3.44. Then, we have

$$\begin{aligned} \int_0^{T_0} r(t) \hat{u}_i^*(t) dt &= \sum_{n=1}^N \hat{u}_{in}^* \int_0^{T_0} r(t) f_n^*(t) dt \\ &= \sum_{n=1}^N \hat{u}_{in}^* r_n \quad (i = 1, 2, \dots, I_m) \end{aligned}$$



**FIGURE 3.40** Optimum receiver implementation. (a) Correlator receiver and (b) matched filter receiver.

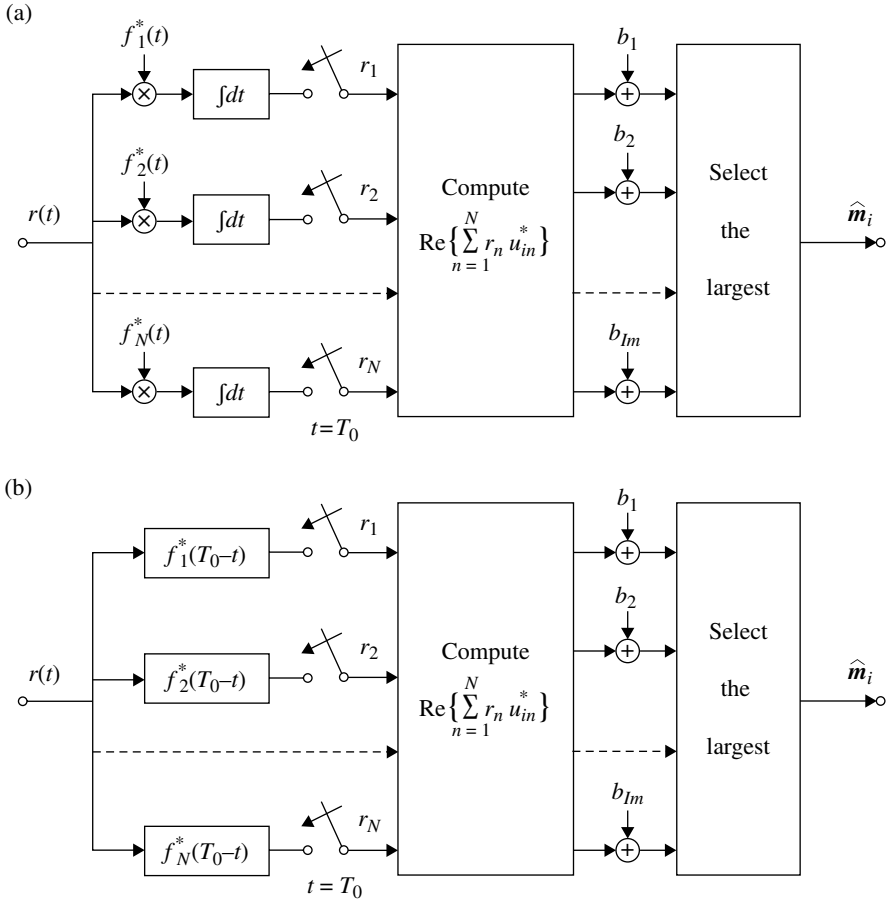
This relation suggests another configuration of the optimum receiver as shown in Figure 3.41.

### 3.3.8 The Maximum-Likelihood Receiver and the Viterbi Algorithm

The number of calculations to find the optimum candidate message  $\hat{m}_i$  increases exponentially with the message length,  $N$ , and is proportional to  $I_m = L^N$ , where  $L$  is the number of digital signal states. A brute force method is ineffective for long messages. Instead, the Viterbi algorithm or dynamic programming is used to decrease drastically the number of calculations.

In practice, it is difficult to know  $P(\mathbf{m}_i)$ , the probability of the occurrence of the message  $\mathbf{m}_i$ , so we may assign equal probabilities to all the messages. The optimum receiver under this condition is called the maximum-likelihood receiver.





**FIGURE 3.41** Other configuration of the optimum receiver. (a) Correlator receiver and (b) matched filter receiver.

The likelihood is maximized by finding out the candidate message  $\hat{m}_i$  or equivalently the candidate signal  $u_i(t)$  that minimizes the integral:

$$V_i = \int_0^{T_0} |r(t) - \hat{u}_i(t)|^2 dt$$

Minimizing this integral corresponds to maximizing Equation 3.48. Expanding the function under the integral sign, the objective function to be maximized can be expressed as follows:

$$V_i = -2 \text{Re} \left[ \int_0^{T_0} r(t) \hat{u}_i^*(t) dt \right] + \int_0^{T_0} |\hat{u}_i(t)|^2 dt \quad (3.49)$$

where the term  $\int_0^{T_0} |r(t)|^2 dt$  is ignored since it is irrelevant to the decision-making. Using Equation 3.39, Equations 3.49 can be rewritten as:

$$V_i = -2 \operatorname{Re} \left[ \sum_{n=1}^N \hat{a}_{in}^* r_n \right] + \sum_{n=1}^N \sum_{m=1}^N \hat{a}_{in} \hat{a}_{im}^* q_{n-m} \quad (3.50)$$

where

$$r_n = \int_0^{T_0} r(t) g^*(t - (n-1)T_s) dt \quad (n = 1, 2, \dots, N)$$

and

$$q_{n-m} = \int_0^{T_0} g(t - nT_s) g^*(t - mT_s) dt \quad (3.51)$$

For simplicity, we omit the suffix  $i$  that represents the  $i$ th message. We assume the following:

$$q_{n-m} = 0 \quad (|n-m| > M) \quad (3.52)$$

This assumption means that the symbols  $a_m$  and  $a_n$  never interfere with each other for  $|n-m| > M$ . Equation 3.50 can be rewritten (omitting the suffix  $i$ ) as:

$$V = \sum_{n=1}^{j-1} K_n + K_j + L_{N-j} \quad (j = 1, 2, 3, \dots, N) \quad (3.53)$$

where

$$\begin{aligned} K_n &= -2 \operatorname{Re} [\hat{a}_n^* r_n] + 2 \operatorname{Re} \left[ \hat{a}_n \sum_{m=n+1}^{M+n} \hat{a}_m^* q_{n-m} \right] + |\hat{a}_n|^2 q_0, \\ L_{N-j} &= -2 \operatorname{Re} \left[ \sum_{n=j+1}^N \hat{a}_n^* r_n \right] + \sum_{n=j+1}^N \sum_{m=j+1}^N \hat{a}_n \hat{a}_m^* q_{n-m} \end{aligned} \quad (3.54)$$

We applied Equation 3.52 to the second term of Equation 3.54 and  $a_i = 0$  for  $i > N$  is understood. The reason why we express the objective function  $V$  (metric) as Equation 3.53 will be understood through description in the following.

Upon receiving the signal  $r_i$ , we start the calculation of the objective function  $V$  (metric): the values of  $K_1$  are calculated for  $L^{M+1}$  candidate message sequences  $(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_{M+1})$ . We can divide these  $L^{M+1}$  sequences into  $L^M$  groups, corresponding to the possible choices of  $(\hat{a}_2, \dots, \hat{a}_{M+1})$ . Each group consists of  $L$  sequences that are different only in the value of  $\hat{a}_1$ . For each group, we choose the state of  $\hat{a}_1$  that minimizes  $K_1$ , and the other  $L-1$  states of  $\hat{a}_1$  are discarded. The chosen  $\hat{a}_1$  and the corresponding minimum  $K_1$  value, denoted as  $K_{\min}$ , are stored for each sequence group. Let  $\hat{\sigma}_n$  denote the candidate message sequence  $(\hat{a}_{n+1}, \hat{a}_{n+2}, \dots, \hat{a}_{n+M})$ . Thus,

$K_{1\min} = \min_{\hat{\sigma}_1} [K_1(r_1; \hat{a}_1, \hat{\sigma}_1)]$ . Discarding the  $L-1$  states of  $\hat{a}_1$  for each  $\hat{\sigma}_1$  is justified by the fact that they could not be a candidate later. This is true since their contribution to the metric for the signal sequences is limited within  $\hat{\sigma}_1$ .

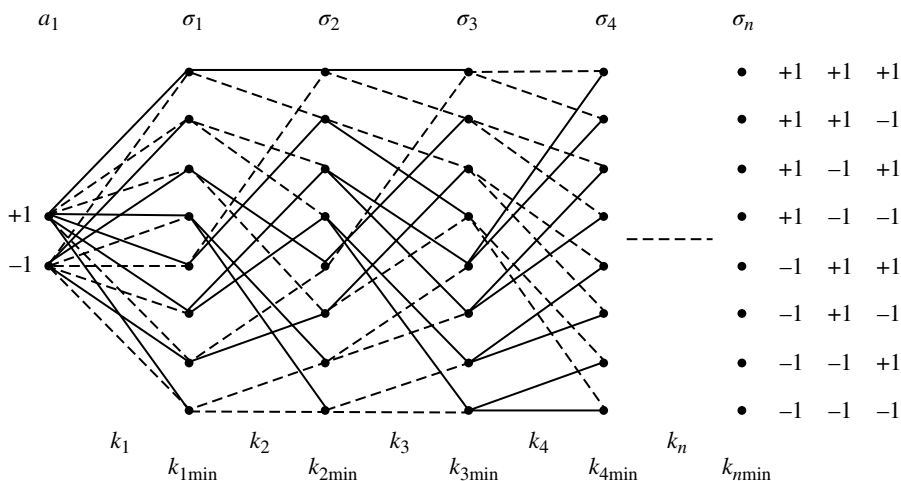
At this stage, if all the sequence groups  $\hat{\sigma}_1$  have the same value, then we decide on  $\hat{a}_1$ . Otherwise, the decision on  $\hat{a}_1$  is deferred to a later time.

Receiving the signal  $r_2$ , the  $K_2$  values called branch metrics are calculated for  $L$  candidates  $\hat{a}_2$ . The  $L^M$  sequences  $\hat{\sigma}_2$  are calculated in the same way as  $\hat{a}_1$  and  $\hat{\sigma}_1$  in the previous step of the algorithm. The path metric, that is,  $K_{2\min} = \min_{\hat{\sigma}_2} [K_{1\min} + K_2(r_2; \hat{a}_2, \hat{\sigma}_2)]$ , is selected for each  $\hat{\sigma}_2$  with the corresponding  $\hat{a}_2$  and the others are discarded. The selected  $\hat{a}_2$  and the path metrics  $K_{2\min}$  are stored for each  $\hat{\sigma}_2$ . If all the sequences  $\hat{\sigma}_2$  have the same  $\hat{a}_1$  or  $(\hat{a}_1, \hat{a}_2)$  value, a decision is made on these values. Otherwise, the decision-making is deferred to a later time.

Receiving a new signal  $r_n$ , we proceed with the calculation and the selection in the same way: we have  $L^{M+1}$  possible message sequences  $(\hat{a}_n, \hat{a}_{n+1}, \dots, \hat{a}_{n+M})$  which are the continuations of the  $L^M$  surviving sequences of the previous stage of the process. We divide these sequences into  $L^M$  groups, each group containing  $L$  sequences with the same  $\hat{\sigma}_n$  value, differing only in  $\hat{a}_n$ , and calculate the branch metric  $K_n$ . From each group, we select  $\hat{a}_n$  that minimizes  $[K_{(n-1)\min} + K_n(r_n; \hat{a}_n, \hat{\sigma}_n)]$ , discarding the other  $L-1$  sequences. The selected  $\hat{a}_n$  together with the precursors  $(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_{n-1})$  and the branch metric  $K_{n\min}$  is stored. At the end of the procedure, we can make decision on a sequence that shows the minimum path metric.

In each stage, we compute  $K_j$  for  $L^{M+1}$  sequences. It means  $NL^{M+1}$  computations for detecting  $N$  symbols.

The lattice diagram of the Viterbi algorithm is shown in Figure 3.42 for an example ( $M=3$ ). The succession of the state  $\hat{\sigma}_n$  has one-to-one correspondence to the symbol



**FIGURE 3.42** Lattice diagram of the Viterbi algorithm. Solid lines and dashed lines show the surviving and the discarded paths, respectively.

sequence of  $\{a_n\}$ . Solid line shows the survived path and dashed line represents the discarded path. When all the survived paths derive from the same state  $\hat{\sigma}_n$ , that is, merging occurs, we can make a decision on that symbol  $\hat{a}_n$  and its precursors. The merging occurs stochastically depending on the signal sequence and noise. If merging does not occur, the decision-making is made at the end of the received signal. The total calculations on the metrics is  $NL^{M+1}$ , in contrast to  $L^N$  with the brute force method, for  $N \gg 1$ . Selecting one candidate from the  $L$  states of a symbol for a sequence  $\hat{\sigma}_n$  at every stage of process and discarding the others become the key of the Viterbi algorithm in reducing the number of calculations.

### 3.3.9 The Optimum Receiver for Signals without Intersymbol Interference

Consider signals for which the matched filter output shows no intersymbol interference. Since the output of a matched filter is equivalent to that of a correlation receiver at the sampling instants, the signal  $g(t)$  fulfills  $q_{n-m} = 0$  for  $n \neq m$  (Eqs. 3.51 and 3.33). This means that there is no correlation between the sampled signals  $r_n$ .

The objective function  $V_i$  given by Equation 3.50 for the optimum receiver becomes

$$V_i = -2\text{Re} \left[ \sum_{n=1}^N \hat{a}_m^* r_n \right] + \sum_{n=1}^N |\hat{a}_{in}|^2 q_0 \quad (i = 1, 2, \dots, I_m) \quad (3.55)$$

where suffix  $i$  has been used again. This result shows that a decision-making on a symbol  $\hat{a}_{in}$  never affects decision-making on other symbols. Thus, the symbol-by-symbol decision becomes the optimum (equal  $P(m_i)$  is assumed) because there is no correlation between the different symbols at the sampled output of the matched filter or correlator. NRZ signaling and transmit pulse shaping with a filter that has a transfer function of the square root of the Nyquist-I characteristics fall into this case.

Instead of examining the objective function of Equation 3.55, we can adopt threshold detection as follows. The threshold detection is made by classifying the received signals  $r_n$  into one of the regions whose boundaries are the threshold values (Fig. 3.43). The detection algorithm is equivalent to determining a level, that is the nearest to  $r_n$ , as

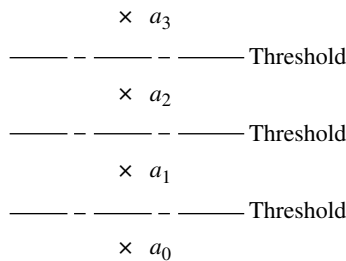


FIGURE 3.43 Threshold detection.

$$\begin{aligned}
 d_m &= \min_{\hat{a}_m} \int_0^{T_0} \left[ r(t) g^*(t - (n-1)T) - \hat{a}_m |g(t - (n-1)T)|^2 \right] dt \\
 &= \min_{\hat{a}_m} |r_n - \hat{a}_m q_0|
 \end{aligned}$$

Taking the square of  $d_m$ , we have

$$d_m^2 = \min_{\hat{a}_m} \left\{ |r_n|^2 - 2 \operatorname{Re} [\hat{a}_m^* r_n q_0] + |\hat{a}_m|^2 q_0^2 \right\}$$

The term  $|r_n|^2$  is irrelevant to decision-making. The second and the third terms are the same as those in Equation 3.55 except for the constant  $q_0$ .

The thresholds must be drawn at the midpoints of the neighboring signal levels  $a_m q_0$  (when  $P(m_i)$  are unknown).

### 3.4 SYNCHRONIZATION

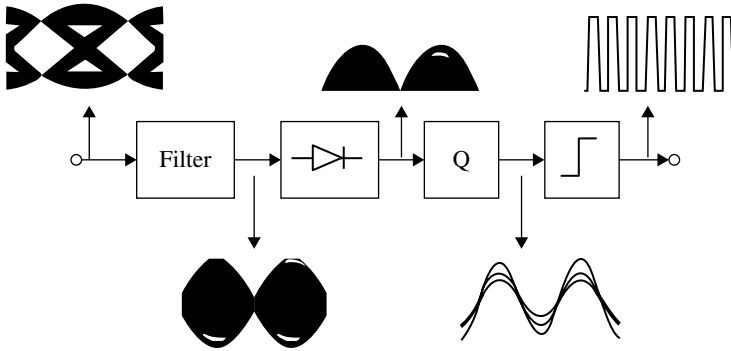
Receive symbol timing and frame timing must be synchronized with the transmitted signal. This section describes how to establish this synchronization.

#### 3.4.1 Symbol Timing Recovery

We do not send permanently the symbol timing signals along with the information signals, since it would decrease the capacity of transmission of the information signals. Instead, we send a short signal only at the start of the transmission for the purpose of aiding the receiver in generating the timing signal. After receiving the timing signal, two strategies are followed to maintain the timing signal: one is to use a sufficiently stable oscillator and the other is to continue extracting the timing signal from the information signals. The former method can be rather easily adopted for low-speed transmission, since a timing error owing to the frequency instability of the oscillator becomes small relative to the symbol duration.

The latter method is used widely. The principle of this method is based on detecting the change in the signal level or polarity, which occurs according to the change in data signals. Although the change in signal levels takes place at random, its average timing is precise since it is synchronized to the transmit timing generator. Figure 3.44 shows a block diagram of the symbol timing regenerator. The demodulated signal is applied to a filter that emphasizes the signal level change. The filtered signal is full-wave rectified. The rectifier is necessary to regenerate the timing signal components, since the level change occurs at random in the opposite directions, resulting in the cancellation of the timing signal.

The tank circuit is a resonator that is tuned to the clock signal frequency. It is intended to suppress the jitter of the regenerated clock timing signal. The  $Q$ -value of the resonator is chosen by a compromise between the performance for low jitter and short rise-up time; a high  $Q$ -value decreases the jitter but increases the rise-up time.



**FIGURE 3.44** Symbol timing regeneration circuit and its waveforms.

When the noise is absent, we can design the pre-rectification filter so that jitter never occurs. Jitter never occurs if the filtered waveform crosses the zero level at the exact midpoint of the symbol times. That is, the impulse response  $h(t)$  of the system consisting of the filter with other filters in the channel must meet the condition:

$$h(t) \sum_{n=-\infty}^{\infty} \delta \left[ \left\{ t - \left( n + \frac{1}{2} T \right) \right\} \right] = 0$$

where  $T$  denotes symbol duration. Taking the Fourier transform of the above equation, and proceeding in a manner as in Section 3.1.2, we get

$$H \left( \frac{\omega_0}{2} + x \right) = H^* \left( \frac{\omega_0}{2} - x \right) \quad (3.56)$$

where  $\omega_0 = 2\pi/T$ , and we assumed that the bandwidth is limited within  $0 \sim \omega_0$ . The left and right sides of Equation 3.56 must be conjugate mirror images of each other with respect to half of the symbol frequency. The computer simulated waveforms in the system with pre-rectification filter satisfying Equation 3.56 are shown in Figure 3.44.

We could not conclude that the pre-rectification filter with Equation 3.56 is still the best with the presence of noise. The optimum clock signal regeneration may depend on the signal to noise power ratio at the demodulator output. There are many other clock signal regeneration systems but are beyond the scope of this book.

### 3.4.2 Frame Synchronization

Digital signals are transmitted collectively in a unit of frame, as it is shown in Figure 3.45. A framed signal consists of information signals and some extra signals, one of which is the frame synchronization (SYNCH) word. The SYNCH word is used for a receiver to find out the starting point of the information signals. Supplied with the received data signal and the regenerated symbol timing signal, the receiver uses a correlator or a matched filter to detect the SYNCH word. For this purpose, a code with a sharp correlation function is desirable. A failure in the SYNCH word detection due to

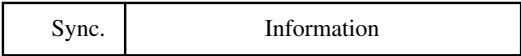


FIGURE 3.45 Framed signal format.

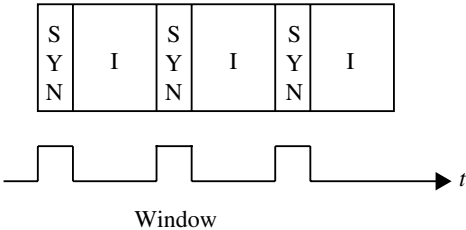


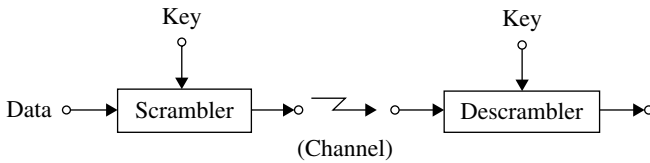
FIGURE 3.46 Frame synchronization for periodically transmitted signals.

symbol errors makes it impossible to receive all the information signals. There are two modes of failure: (i) the receiver misses the detection of the SYNCH word and (ii) the receiver detects an erroneous SYNCH word. We can control the failures by allowing some degree of mismatch between the received and local SYNCH word. By allowing higher degree of mismatch, the former failures can be decreased. However, it decreases the latter failures. Thus, we must have a trade-off between the two failures by taking symbol error probability into consideration. A longer SYNCH word can decrease both failures. However, it decreases the efficiency of the channel.

For a system where the framed signals are periodically transmitted, we can use a shorter SYNCH word. This result is because the second type of failure can be decreased by introducing a periodical window for detecting the SYNCH word in synchronism with the reception of the framed signals, illustrated in Figure 3.46. Since the SYNCH word is short, it is dangerous to decide the SYNCH timing by only one-time detection of the SYNCH word. Therefore, the receiver makes decision on the timing only when the SYNCH words are successively detected several times, after the receiver is in a locked mode. The randomness of data signals justifies the approach of the method. We must pay attention to maintaining the frame timing and fast detecting the loss of synchronization. The receiver recognizes the lack of synchronization after failing to detect the SYNCH word several times successively. Then, the receiver changes into a hunting mode to reestablish the synchronization. The frame synchronizing circuit opens the window for the SYNCH word correlator to output the frame timing signals. This operation was previously explained. The principle of this frame synchronizing method is based on the “flywheel effect,” which is produced by successive detections of the SYNCH word.

3.5 SCRAMBLING

In contrast to the analog signal scrambling, the scrambling of digital signals is easily carried out by using logic circuits or a processor. A purpose of scrambling is to secure the digital transmission. For example, consider a dc cut-off channel; the digital signal



**FIGURE 3.47** Signal transmission with scrambling.

consisting of a constant sequence of “0” or “1” data will not be transmitted, since the signal becomes a dc signal. Scrambling prevents such situation by randomizing the transmitted data. A long sequence of “0” or “1” data often occurs in data terminals. Another case is a channel with the automatic equalizer that requires random data for a stable operation. Standard scrambling is recommended in data communication systems.

Another purpose of scrambling is to protect the signal transmission from eavesdropping by a third party. The degree of the protection can be made sufficiently high using sophisticated digital scrambling methods.

Figure 3.47 shows the block diagram of a communication system with scrambling. The scrambling algorithm may or may not be open to everyone. Even if the scrambling algorithm and hence the descrambling algorithm is open, the key must be known for descrambling. Thus, the number of keys must be large enough for protection against a trial-error or a systematic attack. The data encryption standard (DES), described later, offers  $2^{56}$  ( $\approx 7 \times 10^{16}$ ) different keys. Key management, such as key distribution, becomes an important issue.

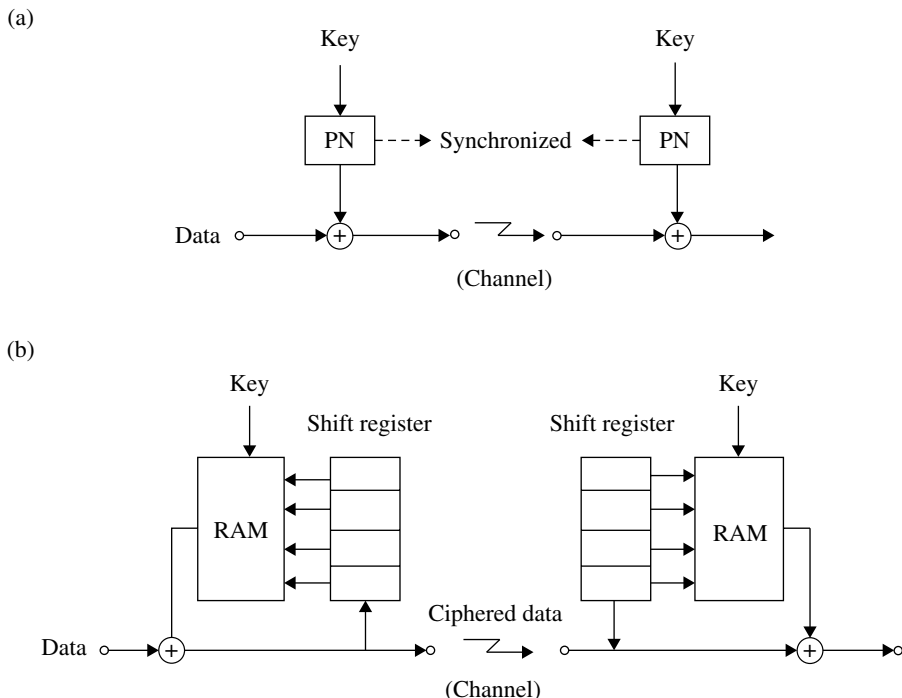
There are several different ways of scrambling: addition of pseudonoise (PN), permutation, and transformation. The addition of pseudonoise literally means that random data is added to the transmitted signal. In the permutation scrambling, the unit of the permutation is a bit or a block of bits (word) and it is performed within a frame or over several frames. The transformation is performed in a unit of a word.

For synchronization between scrambler and descrambler, there are synchronized and self-synchronized systems. Figure 3.48a and b shows the synchronized and the self-synchronized scrambling system, respectively.

The two methods of frame synchronization described before can be applied to synchronize the descrambler to the scrambler. The method where the SYNCH word is transmitted only at the start of the transmission can enhance the security of communication since a third party who missed the SYNCH word will not be able to synchronize his descrambler to the scrambled signal.

The operation of the self-organized scrambling system is explained as follows. The input data is scrambled (ciphered) by the output of a RAM (random access memory). The ciphered data is transmitted into the channel and is fed at the same time into a shift register. The data contained in the shift register determines a RAM address. The RAM outputs random data corresponding to this address value. The output of the RAM is possibly a highly nonlinear function of the input data. Thus, the scrambler can be called a nonlinear feedback circuit. The received data is fed into the shift register and is descrambled by the output of the RAM. The contents of the RAM in the scrambler and descrambler are the same.

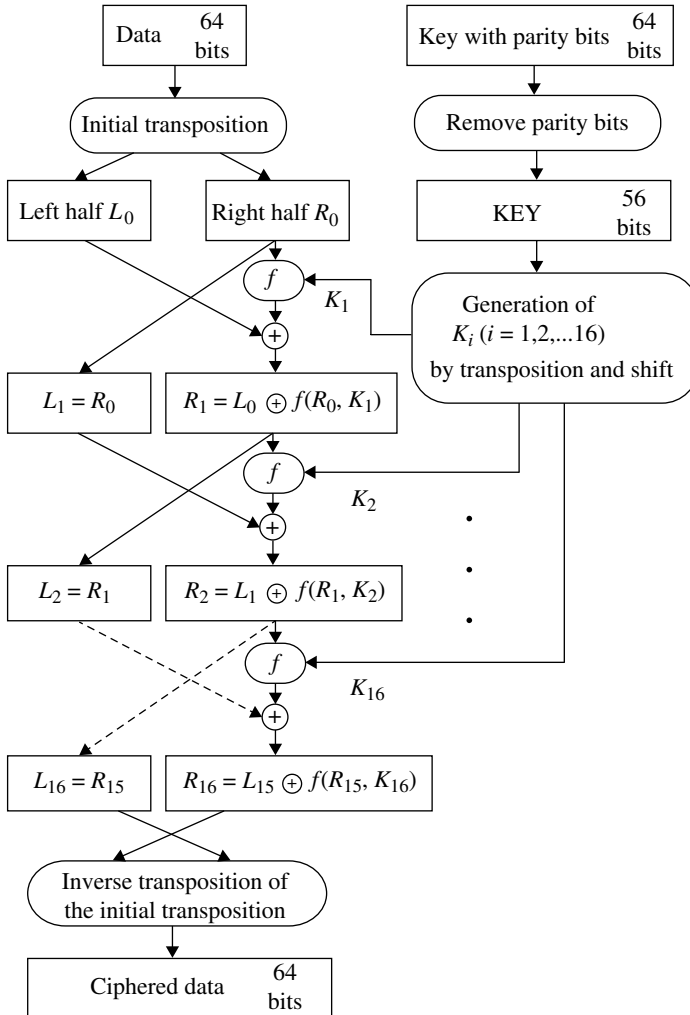




**FIGURE 3.48** Scrambling/descrambling system. (a) Synchronized system and (b) self-synchronized system.

The synchronization of the descrambler to the scrambler is automatically established after the first transmitted signal reaches the end of the shift register: the addresses are always the same for the scrambler and the descrambler RAMs, as long as the transmission error is absent. The penalty we must pay for using the self-synchronized scrambling system is the error propagation effect. A transmission error in the data affects the descrambler while the data errors are in the shift register. Taking into consideration that the RAM outputs “0” or “1” with equal probability for random data, we have on average error length of  $N/2$  at the scrambler output, where  $N$  is the length of the shift register. In the synchronized scrambling system, we have never such error propagation effect. This effect results in deteriorating the sensitivity of the receiver.

The well-known scrambling algorithm DES [13] is shown in Figure 3.49. The data are processed in a unit of 64 bits. The input data is subjected to an initial transposition. The transposed data is divided into left- and right-hand sides. The right-hand side data serves as the left-hand side data of the subsequent stage and is substituted using the subkeys of 48 bits. The substituted data are then added to the left-hand side data so as to generate the right-hand side data. This process is repeated 16 times with subkeys  $K_i$  ( $i = 1, 2, \dots, 16$ ). The final data is transposed in an opposite way of the initial transposition. The key consists of 56 bits and 8 parity bits. The subkeys are generated by substituting and shifting the key.

**FIGURE 3.49** DES scrambler.

To enhance the security of DES, advanced encryption standard (AES) was standardized in 2001. The data block length became 128 bits, and three key lengths of 128, 192, or 256 bits are used.

### 3.6 PUBLIC KEY CRYPTOSYSTEM

The scrambling or cryptography described so far requires secret keys for both encryption and decryption at transmitter and receiver sides, respectively. The public key cryptosystem uses a secret key only for decryption. The well-known RSA (invented by R. Rivest, A. Shamir, and L. Adleman) cryptosystem is described in the following text.

We divide a message into blocks of a given length and express a block with an integer  $M$ . At a transmitter,  $M$  is encrypted in the given manner using an encryption key that is open to public corresponding to each receiver and sent to the receiver. The receiver decrypts the encrypted data to get  $M$  using a secret key that is known only to the receiver. For a cryptosystem, it is desirable to have the following criteria: (i) robustness against attack, (ii) a less complexity in encryption and decryption procedure, and (iii) slight increase in the encrypted data bit length. The RSA meets these conditions by difficulty of factoring large integers for (i), power operation for (ii), and modulo operation for (iii). We use mathematical expression to explain the encryption and decryption processes and their principles. We prepare two sufficiently large prime numbers  $p$  and  $q$ . Then we propose a number  $n$  as follows:

$$n = pq \quad (3.57)$$

where  $n > M$ . For a larger  $n$ , it is difficult to get  $p$  and  $q$  from  $n$ . We make  $z = (p-1)(q-1)$  and choose an arbitral integer  $e$  that is relatively prime (i.e., the greatest common divisor is 1) with  $z$ . Then we find  $d$  such that

$$ed = (p-1)(q-1)k + 1 \quad (k : \text{integer})$$

or

$$ed \pmod{(p-1)(q-1)} = 1$$

Encryption is done with integer  $e$  and  $n$ , which are open to public, as

$$c = M^e \pmod{n}$$

Decryption is carried out with the secret key number  $d$  as

$$\begin{aligned} r &= c^d \pmod{n} \\ &= M \end{aligned}$$

The principle of the decryption to get the transmit message  $M$  owes to properties of integers as follows.

$$\begin{aligned} r &= \left( M^e \pmod{n} \right)^d \pmod{n} \\ &= M^{ed} \pmod{n} \end{aligned}$$

where the following relation is used:

$$ab \pmod{n} = a \pmod{n} b \pmod{n} \pmod{n}$$

and

$$\left( a \pmod{n} \right)^e = a^e \pmod{n}$$

The next step is based on Fermat's little theorem and the Chinese remainder theorem (Appendix 3.A).

$$\begin{aligned}
 M^{ed} \pmod{n} &= M^{(p-1)(q-1)k+1} \pmod{n} \quad (ed = (p-1)(q-1)k+1) \\
 &= M \cdot M^{(p-1)(q-1)k} \pmod{n} \\
 &= M
 \end{aligned}$$

The public key cryptosystem is used for digital signature to authenticate the message sender and is as follows: The sender encrypts his message with his secret key. The receiver decrypts the received message with the public key assigned for the sender. When the secret and public keys correctly correspond to each other, the receiver has a meaningful text; otherwise the decrypted message is meaningless. The principles of this authentication are based on the capability of order change between encryption and decryption.

### 3.7 MULTIPLEXING AND MULTIPLE ACCESS

The multiplexed transmission, where multiple signals are sent on a given transmission medium, becomes important from system design's viewpoint.

In a multiplexed transmission system, efficient use of spectrum and least interference between signals should be taken into consideration. To get an interference-free transmission, we use a set of signals that are orthogonal with each other (Section 2.1.3). The frequency-division multiplexing (FDM) where signal spectra never overlap has been used for analog transmission systems, since it requires simple signal processing with analog circuits.

The time-division multiplexing (TDM), where signals are multiplexed in time domain, appeared for digital transmission systems, since digital signal is rather easy to be processed for such as time delaying, and separation in time domain, which are required in the TDM. One of the merits of digital system is its stability in processing digital signals with digital circuits.

The code-division multiplexing (CDM) uses orthogonal signals that overlap with each other in time- as well as frequency domain, such as the Walsh codes (Section 2.1.3). This system appeared most lately, since new techniques should be developed to overcome interference between the CDM signals: the interference due to impairment in a transmission media is severe in CDM compared with FDM and TDM. In recent years, it has been applied in cellular systems due to its high-spectrum efficiency, which are realized through new techniques (Section 9.7.2) and progress in digital signal processing performance.

In a mobile radio communication system, multiplexed transmission is done in the downlink (from base station to multiple mobile stations). In the uplink (from mobile to base), multiple users make access to and share the radio channel. To share the channel, the aforementioned multiplexing is used. In this case we call them as FDMA (frequency-division multiple-access), TDMA (time-division multiple-access), and CDMA (code-division multiple-access) emphasizing the multiple access to the channel.

The peak to average power ratio (PAPR) of transmit signal becomes a practical issue in a multiplexed (multi-access) system (Section 6.5.1). A lower PAPR is better in view of transmit power amplification (Section 8.6.3). Required average transmit power for receive signal quality (i.e., error rates) never depends on multiplexing method in principle. However, the peak power depends strongly on the multiplexing method: a peak power becomes high proportionally to the number of multiplexed signal when the signals are added in phase at the same time. Therefore, TDM shows advantage over FDM and CDM, since its signals are never added at the same time. Contrary to this, the situation reverses in the uplink. If FDMA or CDMA mobile terminal never uses a multiplexed transmission, the PAPR value is low, which is given with its modulation method. TDM, however, shows a high PAPR value proportionally to the number of time-division multiplexing.

A multiplexed transmission named MIMO system (Section 7.2) which uses multiple transmit and receive antennas without any orthogonal signals is sometimes called SDM(A) (space-division multiplexing (multiple access)). It is assumed that transmit signals propagate in independent paths, and if the paths are dependent with each other, the multiplexed transmission becomes impossible.

### 3.8 THE CHANNEL CAPACITY

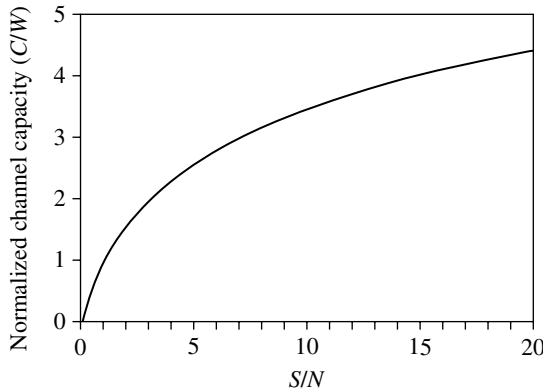
In order for error rates in a given digital transmission system to be infinitesimally zero, the signal to noise power ratio ( $S/N$ ) should become infinitely large as discussed in Section 3.3. The noise power is expressed as  $N = N_0 W$  where  $N_0/2$  is the both-side noise power density and  $W$  is the equivalent noise bandwidth (Section 2.2.1). Decreasing the noise power density is artificially difficult in contrast to the signal power  $S$  and the bandwidth  $W$ . For a given  $N_0$ , we should increase  $S$  infinitely or make  $W$  infinitely narrow resulting in a zero approaching transmission speed. C.E. Shannon showed that this is not necessarily true but we can get a infinitesimally small error rates with a given  $S/N$  under a limit of transmission speed (capacity) and is as follows [14]:

$$C = W \log_2 (1 + S/N) \quad (\text{bps : bits/s}) \quad (3.58)$$

where additive white Gaussian noise is assumed. For the signal, its level obeys Gaussian distribution and it is band-limited in  $W$ . Such a signal or code is possible only in an infinitesimal way, since the Gaussian distribution requires an infinite distribution of the signal level. Toward the channel limit, many practical approaches have been investigated with introduction of error-correcting codes. In recent years, some methods such as the turbo codes and LDPC systems are realized to attain the channel capacity under a practically low error rates (Sections 7.4.5 and 7.4.6).

We will investigate what Equation 3.58 means in the following. Dividing both side of this equation by  $W$ , we have the channel efficiency defined by channel speed per unit frequency bandwidth as

$$\frac{C}{W} = \log_2 (1 + S/N) \quad (\text{bps/Hz})$$



**FIGURE 3.50** Channel capacity versus signal to noise power ratio.

This relation is shown in Figure 3.50. We get  $C \cong WS/N \log_2 e = S/N_0 \log_2 e$  for a region of  $S/N \ll 1$ . The capacity  $C$  depends on the signal power  $S$  and is independent with the bandwidth  $W$  in this power limit region. In a region of  $S/N \gg 1$ , we have  $C \cong W \log_2 S/N = W \log_2 (S/N_0 W)$ .

The variation in  $C$  scarcely depends on that of  $S$  and is almost proportional to that of  $W$  in this bandwidth limit region.

We consider the meaning of the channel capacity formula from another point of view. Therefore, Equation 3.55 is modified as follows:

$$C = \frac{\eta}{2^\eta - 1} \frac{S}{N_0} \quad (3.59)$$

where we put  $C/W = \eta$ .  $\eta$  is the channel efficiency. The channel capacity is proportional to  $S/N_0$  and its proportional coefficient is given as a function of  $\eta$ . The above equation is plotted in Figure 3.51. The upper limit of the proportional coefficient is  $1/\log 2$  ( $\eta = 0$ ) and the lower limit is zero ( $\eta = \infty$ ).

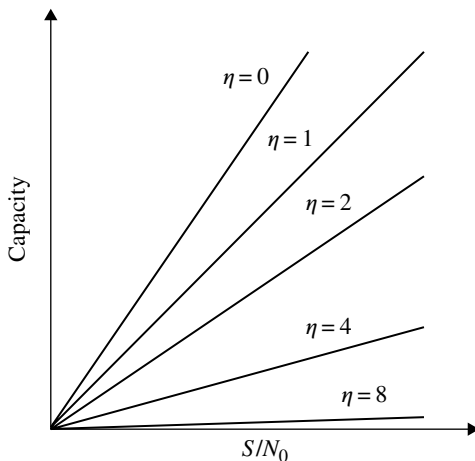
If we let an actual transmission speed as  $R$ , we have  $S = E_b R$  ( $E_b$ : energy per bit) and  $\frac{S}{N} = \frac{E_b}{N_0} \frac{R}{W}$ .

Since the actual transmission speed is less than the channel capacity  $C$ , we have the inequality as

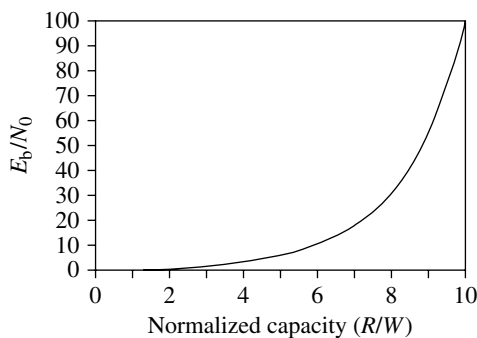
$$R \leq C = W \log_2 \left( 1 + \frac{E_b}{N_0} \frac{R}{W} \right)$$

Letting  $\eta = R/W$  (channel efficiency (bps/Hz)), from the above equation, we have the following:

$$\frac{E_b}{N_0} \geq \frac{2^\eta - 1}{\eta} \quad (3.60)$$



**FIGURE 3.51** Channel capacity versus signal power to noise power density ratio.



**FIGURE 3.52**  $E_b/N_0$  versus normalized channel capacity.

The limit of this relation is shown in Figure 3.52. Let us consider a case when  $N_0$  and  $R$  are fixed. Here,  $\eta$  is strongly dependent on  $E_b$  when  $\eta \ll 1$ . From this, it is known that the required bandwidth  $W$  is greatly decreased with a small increase in  $E_b$  in this region. On the other hand, in the region of  $\eta \gg 1$ , we can decrease  $E_b$  and therefore  $S$  with a small decrease of  $\eta$  and, therefore, with a small increase of  $W$ .

### APPENDIX 3.A FERMAT'S THEOREM AND THE CHINESE REMAINDER THEOREM

#### Fermat's Theorem

For a prime number  $p$  and an integer  $M$  that is relatively prime with  $p$ , Fermat's theorem states

$$M^{p-1} \pmod{p} = 1 \quad (3.A.1)$$

**Proof**

Since  $M \pmod{p} \neq 0$ , we know  $Mi \pmod{p} \neq Mj \pmod{p} (i \neq j), (i, j = 1, 2, \dots, p-1)$ . Consider sets of integers  $\{i\}$  and  $\{Mi \pmod{p}\} (i = 1, 2, \dots, p-1)$ , then  $\{i\}$  and  $\{Mi \pmod{p}\} (i = 1, 2, \dots, p-1)$  are congruent in different order and therefore we have

$$\prod_{i=1}^{p-1} i \pmod{p} = \prod_{i=1}^{p-1} Mi \pmod{p} = M^{p-1} \pmod{p} \prod_{i=1}^{p-1} i \pmod{p}$$

Thus, Equation 3.A.1 is obtained.

**The Chinese Remainder Theorem**

Consider different prime numbers  $p$  and  $q$ , and let  $n=pq$ ,

$$X_1 = X \pmod{p} \text{ and } X_2 = X \pmod{q} \quad (3.A.2)$$

for a positive integer  $X$ , then the Chinese remainder theorem states:

$$X \pmod{n} = X_1 q_p^{-1} q + X_2 p_q^{-1} p \pmod{n} \quad (3.A.3)$$

where  $q_p^{-1}$  and  $p_q^{-1}$  are integers such that  $q_p^{-1} q \pmod{p} = 1, p_q^{-1} p \pmod{q} = 1$ .

**Proof**

Equation 3.A.3 is sufficient for Equation 3.A.2 to hold. A set of  $X_1$  and  $X_2$  corresponds to  $X \pmod{n}$  uniquely; therefore, Equation 3.A.3 is necessary for Equation 3.A.2 to hold.

If we put  $X=1$  in Equations 3.A.2 and 3.A.3, we get  $X_1=X_2=1$  and

$$q_p^{-1} q + p_q^{-1} p \pmod{n} = 1 \quad (3.A.4)$$

Letting  $X=M^{(p-1)(q-1)k}$ , from Fermat's theorem, we have the following:

$$\begin{aligned} X_1 &= X \pmod{p} = \left[ M^{(p-1)} \pmod{p} \right]^{(q-1)k} = 1 \\ X_2 &= X \pmod{q} = \left[ M^{(q-1)} \pmod{q} \right]^{(p-1)k} = 1 \end{aligned}$$

Therefore, we get the desired result and is as follows:

$$M^{(p-1)(q-1)k} \pmod{n} = q_p^{-1} q + p_q^{-1} p \pmod{n} = 1$$

*Example of the Chinese Remainder Theorem:*  $X$  is less than 35 and is given such that  $X \pmod{5} = 2$  and  $X \pmod{7} = 3$ , then what is  $X$ ? Since  $3 \cdot 5 \pmod{7} = 3 \cdot 5 \pmod{7} = 1$ , answer is  $X \pmod{35} = 2 \cdot 3 \cdot 7 + 3 \cdot 3 \cdot 5 \pmod{35} = 17$ .



**REFERENCES**

- [1] W. R. Bennett and J. R. Davey, *Data Transmission*, McGraw-Hill, New York, 1965.
- [2] M. Schwartz, W. R. Bennett, and S. Stein, *Communication Systems and Techniques*, McGraw-Hill, New York, 1966.
- [3] R. W. Lucky, J. Saltz, and E. J. Weldon, Jr., *Principles of Data Communication*, McGraw-Hill, New York, 1968.
- [4] B. P. Lathi, *Modern Digital and Analog Communication Systems*, in *HRW Series in Electrical and Computer Engineering*, Holt, Rinehart and Winston, New York, 1983.
- [5] J. G. Proakis, *Digital Communications*, 3rd ed., McGraw-Hill, New York, 1995.
- [6] J. G. Proakis and M. Salehi, *Communication Systems Engineering*, Prentice-Hall, Englewood Cliffs, 1994.
- [7] E. A. Lee and D. G. Messerschmitt, *Digital Communication*, Kluwer Academic Publishers, Norwell, 1994.
- [8] S. Haykin, *Communication Systems*, 3rd ed., Wiley, New York, 1994.
- [9] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, Wiley, New York, 1965.
- [10] W. B. Davenport, Jr. and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*, IEEE Press, New York, 1987.
- [11] M. Schwartz, *Information Transmission, Modulation, and Noise*, McGraw-Hill, New York, 1980.
- [12] T. J. Richardson, M. A. Shokrollahi and R. L. Urbanke, Design of capacity-approaching irregular low-density parity-check codes, *IEEE Trans Inf Theory*, 47 (2), pp. 619–637, 2001.
- [13] A. S. Tannenbaum, *Computer Networks*, Prentice-Hall, Englewood Cliffs, 1981.
- [14] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1963.

This page intentionally left blank

---

# 4

---

## MOBILE RADIO CHANNELS

In wireless communication, the received signal power  $P_r$  is expressed as:

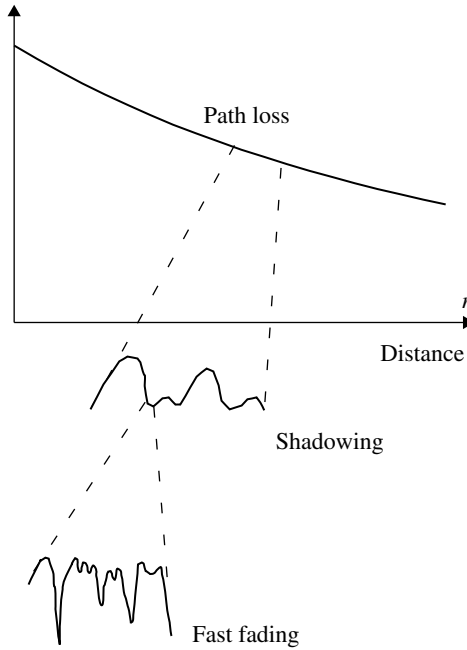
$$P_r = G_r L_c G_t P_t \quad (4.1)$$

where  $P_t$  is the transmit signal power,  $G_t$  the transmit antenna gain,  $G_r$  the receive antenna gain, and  $L_c$  the propagation loss in the channel. In an ideal free space, the propagation loss  $L_c$  is known as the Friis formula and is as follows:

$$L_c = \left( \frac{\lambda}{4\pi d} \right)^2$$

where  $d$  is the distance between the transmit and receive points and  $\lambda$  is the wavelength of the radio wave. This is a simple equation. However, it is rather hard to understand in a physical sense that the wavelength  $\lambda$  term is included. A derivation of the formula with a concept of the antenna aperture cross section is given in Appendix 4.A.

A land mobile radio channel is characterized by out-of-sight communication to/from a moving terminal; it becomes a multipath propagation channel with fast fading. The characteristics of the mobile radio channel introduce new challenges for the designer of mobile communication systems, for example, requirements for isotropic directivity antenna, choice of appropriate carrier frequency, and techniques for stable transmission under the condition of fast fading.



**FIGURE 4.1** Schematic diagram of the propagation loss.

Wave propagation in the multipath channel depends on the actual environment, such as antenna height, profile of buildings, roads, and terrain. Therefore, we must describe mobile radio channels in a statistical way. Wave propagation in a mobile radio channel is characterized by three aspects: path loss, shadowing, and fast fading:

$$L_c = L_p L_s L_f \quad (4.2)$$

where  $L_p$ ,  $L_s$ , and  $L_f$  denote path loss, shadowing loss, and fading loss, respectively (Fig. 4.1). The path loss  $L_p$  is an average propagation loss over some wide areas. It is determined by macroscopic parameters, such as the distance between the transmitter and receiver, the carrier frequency, and land profile. The shadowing loss  $L_s$  represents variation of the propagation loss in a local area (several ten meters). The shadowing is caused by variation of propagation conditions due to buildings, roads, and other obstacles in a relatively small area. The fast fading  $L_f$  is due to the motion of the terminal in a standing wave that consists of many diffracted waves; it represents the microscopic aspect of the channel in an order of the wavelength.

#### 4.1 PATH LOSS

The simplest formula for path loss is

$$L_p = Ar^{-\alpha} \quad (4.3)$$

where  $A$  and  $\alpha$  are propagation constants and  $r$  is the distance between the transmitter and the receiver.  $\alpha$  takes a value of 3–4 in a typical urban area. For predicting the propagation constants, many nomographs are obtained through propagation measurements [1–3]. The Okumura curves [1] are famous for practical use. Based on the Okumura curves, Hata [2] (Copyright © IEEE 1980) presented an empirical formula for prediction of path loss. The results are cited here:

(a) Urban Area

$$L_{\text{pU}} (\text{dB}) = 69.55 + 26.16 \log f_c - 13.82 \log h_b - a(h_m) + (44.9 - 6.55 \log h_b) \log R \quad (4.4)$$

where  $L_{\text{pU}} (\text{dB}) = -10 \log L_{\text{pU}}$  and  $a(h_m)$  is a correction factor for vehicular antenna height and is given as follows:

*Small- and medium-size city:*

$$a(h_m) = (1.1 \log f_c - 0.7) h_m - (1.56 \log f_c - 0.8)$$

*Large city:*

$$a(h_m) = \begin{cases} 8.29 (\log 1.54 h_m)^2 - 1.1 & (f_c \leq 200 \text{ MHz}) \\ 3.2 (\log 11.75 h_m)^2 - 4.97 & (f_c \geq 400 \text{ MHz}) \end{cases}$$

(b) Suburban Area

$$L_{\text{pS}} (\text{dB}) = L_{\text{p}} (\text{dB}; \text{urban area}) - 2 \left\{ \log \left( \frac{f_c}{28} \right) \right\}^2 - 5.4 \quad (4.5)$$

(c) Open Area

$$L_{\text{pO}} (\text{dB}) = L_{\text{p}} (\text{dB}; \text{urban area}) - 4.78 (\log f_c)^2 + 18.33 \log f_c - 40.94 \quad (4.6)$$

where  $f_c$  is the frequency (MHz) (150–1500 MHz)

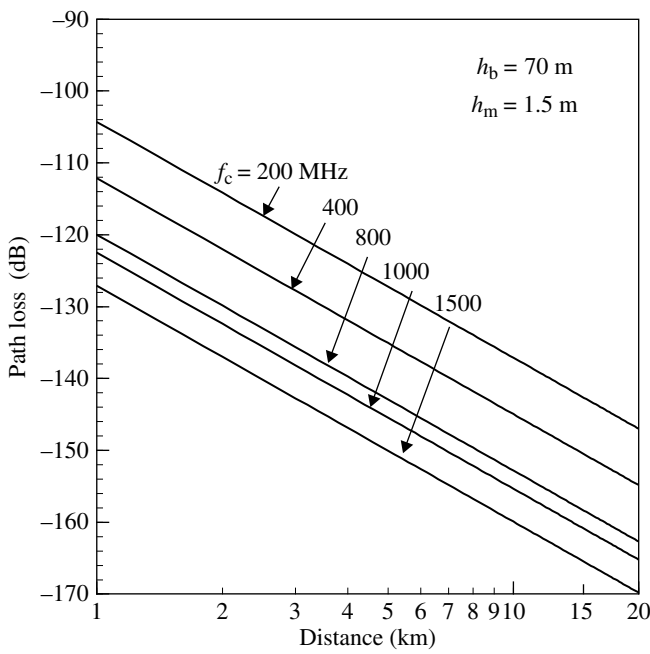
$h_b$  is the base station effective antenna height (m) (30–200 m)

$h_m$  is the vehicular station antenna height (m) (1–10 m)

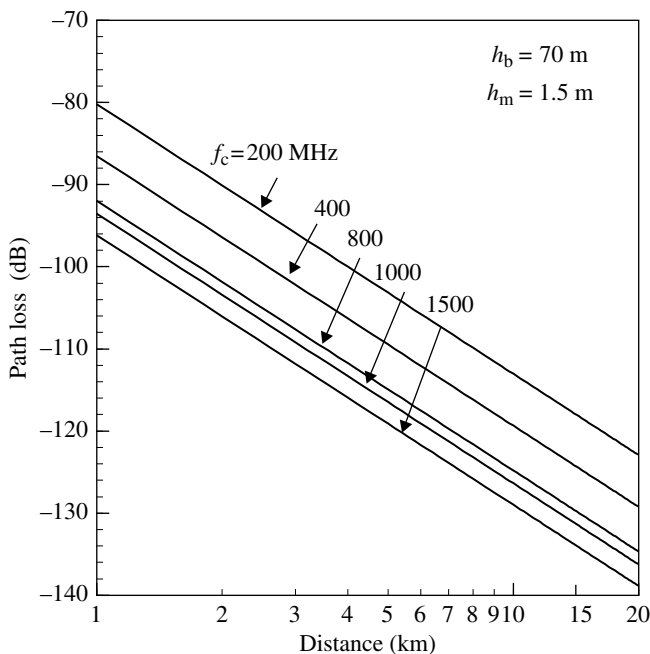
$R$  is the distance (km) (1–20 km)

$\log x$  is the logarithmic function  $\log_{10} x$ .

Path loss characteristics in urban area for small- and medium-size city and for large city are shown in Figures 4.2 and 4.3, respectively. For a large-size city, the path loss is the same as that for small- and medium-size city under  $h_m = 1.5$  m.



**FIGURE 4.2** Path loss vs. distance (small- and medium-size city).



**FIGURE 4.3** Path loss vs. distance (open area).

## 4.2 SHADOWING

Many experiments have shown that the shadowing loss obeys the log-normal distribution. In this case, the probability density function of the received signal level in decibel becomes as follows:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-x_0)^2}{2\sigma^2}} \quad (4.7)$$

where  $x$  is the received power in decibel,  $x_0 = \langle x \rangle$ , and  $\sigma^2$  is the variance in decibel. The averaging of  $x$  is defined over a distance that is longer enough for averaging microscopic variation (several wavelengths). The variance takes values of 4–12 dB depending on the propagation environment.

In the following, we develop a hypothetical argument supporting the log-normal distribution for the received signal level. It is probable that the received signal is subjected to a number of random reflections while propagating from the transmitter to the receiver. Then we can express the received signal level as:

$$P_r = P_0 \prod_{n=1}^{\infty} \Gamma_n \quad (4.8)$$

where  $P_0$  denotes the transmitted power and  $\Gamma_n$  are the random power losses due to reflections and propagation. If we express Equation 4.8 in decibel, we have

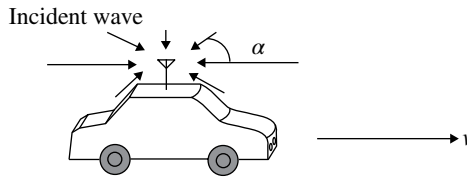
$$10\log_{10} P_r = 10\log_{10} P_0 + \sum_{n=1}^{\infty} 10\log_{10} \Gamma_n$$

The central limit theory (Section 2.2.8) states that the sum of many random values,  $10\log_{10} \Gamma_n$  in our case, obeys the normal distribution. Thus the loss,  $10\log_{10} P_r - 10\log_{10} P_0$ , in decibel obeys the normal distribution.

In actual mobile radio channels, the signal propagates over multipaths instead of one path to the receiver. In this case, the lognormal distribution is also supported, since the addition of several signals with the lognormal distribution is well approximated by a lognormal distribution [4].

## 4.3 FAST FADING

Many random signals that propagate through different signal paths from the transmitter are superposed at the receiver to produce standing waves. When a receiver and/or a transmitter moves in the standing waves, the receiver experiences random variation in the signal level and in the phase and also a Doppler shift. As a consequence of the standing wave characteristic, the minimum distance between signal level drops is one-half of the wavelength; for example, the half wavelength becomes as short as



**FIGURE 4.4** Analytical model for fast fading.

15 cm at a frequency of 1000 MHz. This microscopic deviation of the received signal is called fast fading.

The fast fading phenomena are analyzed in Refs. [3] and [5] with a propagation model where many independent signals reach a moving receiver uniformly from different directions (Fig. 4.4). Here we restrict our argument to cite some important results, solely referring to Ref. [3].

We assume vertically polarized plane waves. The vertical electric field is expressed as follows:

$$E_z = T_c(t) \cos \omega_c t - T_s(t) \sin \omega_c t$$

where  $\omega_c$  is the carrier frequency.  $T_c(t)$  and  $T_s(t)$  are

$$T_c(t) = E_0 \sum_{n=1}^{\infty} C_n \cos(\omega_n t + \phi_n) \quad (4.9)$$

$$T_s(t) = E_0 \sum_{n=1}^{\infty} C_n \sin(\omega_n t + \phi_n) \quad (4.10)$$

where  $N$  is the number of the incident waves,  $E_0 C_n$  is the real amplitude of the electric field ( $C_n$  are normalized, i.e.,  $\sum_{n=1}^N C_n^2 = 1$ ), and  $\phi_n$  are random phases uniformly distributed from 0 to  $2\pi$ .  $\omega_n$  are the Doppler frequencies and are given as

$$\omega_n = \beta v \cos \alpha_n$$

where  $v$  is the velocity,  $\alpha_n$  the incident angle to the direction of the movement,  $\beta$  the propagation constant, which is expressed as,

$$\beta = \frac{2\pi}{\lambda} = \frac{\omega_c}{c}$$

$\lambda$  is the wavelength, and  $c$  is the velocity of light ( $c = 3 \times 10^8$  m/s). The Doppler frequencies vary within  $\pm \beta v$  ( $= \omega_c v/c$ ).

$T_c(t)$  and  $T_s(t)$  are the in-phase and quadrature components of  $E_z$ , respectively. For a modulated signal, they also include the modulating signal. However, a non-modulated carrier signal is treated here for simplicity of argument.  $T_c(t)$  and  $T_s(t)$  become Gaussian random processes for a large  $N$ , as a consequence of the central



limit theorem. We denote the random variables  $T_c(t)$  and  $T_s(t)$  at the time  $t$  by  $T_c$  and  $T_s$ , respectively. They have zero mean, zero cross correlation, and variance:

$$\begin{aligned}\langle T_c \rangle &= \langle T_s \rangle = 0 \\ \langle T_c^2 \rangle &= \langle T_s^2 \rangle = \langle |E_z|^2 \rangle = \frac{E_0^2}{2}, \\ \langle T_c T_s \rangle &= 0\end{aligned}$$

where  $\langle \cdot \rangle$  means ensemble average with respect to  $\alpha_n, \phi_n$ , and  $C_n$ . If we denote  $T_c$  or  $T_s$  by  $x$ , the probability density function of  $x$  is given by

$$p(x) = \frac{1}{\sqrt{2\pi b}} e^{-\frac{x^2}{2b}}$$

where  $b = E_0^2/2$  is the mean power of  $E_z$ . The envelope of  $E_z$  is given by

$$r = (T_c^2 + T_s^2)^{1/2}$$

Referring to the argument in Section 2.2, the envelope  $r$  obeys the Rayleigh distribution and the power  $q = r^2/2$  obeys an exponential distribution. The probability density functions become as follows:

$$p(r) = \frac{r}{b} e^{-\frac{r^2}{2b}} \quad (4.11a)$$

$$p(q) = \frac{1}{b} e^{-\frac{q}{b}} \quad (4.11b)$$

The phase of  $E_z$  is given by

$$\theta = \tan^{-1} \frac{T_s}{T_c}$$

Therefore,  $\theta$  is distributed uniformly, that is, the probability density function becomes

$$p(\theta) = \frac{1}{2\pi} \quad (4.12)$$

#### 4.3.1 RF Power Spectrum Spread due to Fast Fading

Received signal spectrum spreads due to the Doppler effects. The frequency can be expressed as a function of the incident angle and is as follows:

$$f(\alpha) = f_m \cos \alpha + f_c \quad (4.13)$$

where  $f_m = \beta v / 2\pi = v / \lambda$  is the maximum Doppler shift.

We denote incident power included in an interval  $d\alpha$  by  $bp(\alpha)d\alpha$  and the receiving antenna directivity in the horizontal plane by  $G(\alpha)$ . Then the received power within  $d\alpha$  becomes  $bG(\alpha)p(\alpha)d\alpha$ . Using Equation 4.13, this value is equated to the differential variation of the power with frequency as follows:

$$S(f)|df| = b[p(\alpha)G(\alpha) + p(-\alpha)G(-\alpha)]|d\alpha|$$

On the other hand

$$|df| = |-f_m \sin \alpha| |d\alpha| = [f_m^2 - (f - f_c)^2]^{1/2} |d\alpha|$$

Thus

$$S(f) = \begin{cases} b[p(\alpha)G(\alpha) + p(-\alpha)G(-\alpha)][f_m^2 - (f - f_c)^2]^{-1/2} & (0 \leq |f - f_c| \leq f_m) \\ 0 & (\text{otherwise}) \end{cases}$$

where

$$\alpha = \cos^{-1} \frac{f - f_c}{f_m}$$

The power spectrum depends on the antenna directivity  $G(\alpha)$  and the distribution of arrival angles  $p(\alpha)$ . We assume a uniform distribution of arrival angles:

$$p(\alpha) = \frac{1}{2\pi}$$

The antenna directivity depends solely on the kind of the antenna to be used. We assume a vertical whip antenna to sense the  $E_z$  component. The directivity in a horizontal plane becomes:

$$G(\alpha) = 1.5 \quad (4.14)$$

Thus, the received signal power spectrum density is given as follows:

$$S(f) = \begin{cases} \frac{3b}{\omega_m} \left[ 1 - \left( \frac{f - f_c}{f_m} \right)^2 \right]^{-1/2} & (0 \leq |f - f_c| \leq f_m) \\ 0 & (\text{otherwise}) \end{cases} \quad (4.15)$$

where  $\omega_m = 2\pi f_m$ .

### 4.3.2 Correlations Between the In-phase and Quadrature Components

The correlation of the in-phase and quadrature components is given as in Ref. [3]:

$$\langle T_{c1} T_{c2} \rangle = \langle T_{s1} T_{s2} \rangle = g(\tau)$$

$$\langle T_{c1} T_{s2} \rangle = -\langle T_{s1} T_{c2} \rangle = h(\tau)$$

where subscript one and two refer to the times  $t$  and  $t + \tau$ , respectively, and

$$g(\tau) = \int_{f_c - f_m}^{f_c + f_m} S(f) \cos 2\pi (f - f_c) \tau df$$

$$h(\tau) = \int_{f_c - f_m}^{f_c + f_m} S(f) \sin 2\pi (f - f_c) \tau df$$

and  $S(f)$  are the received signal power spectra. Using Equation 4.15,  $g(\tau)$  and  $h(\tau)$  become as:

$$g(\tau) = b_0 J_0(\omega_m \tau), \quad b_0 = 1.5b$$

$$h(\tau) = 0$$

where  $J_0(\cdot)$  is the zeroth-order Bessel function of the first kind. The relation  $b_0 = 1.5b$  appears due to the assumed antenna gain (Eq. 4.14).

### 4.3.3 Correlation of the Envelope

The autocorrelation of the envelope  $r$  is given as

$$R_r(T) \equiv \langle r(t)r(t+\tau) \rangle$$

$$= \frac{\pi}{2} b_0 \left[ 1 + \frac{1}{4} \rho^2(\tau) + \frac{1}{64} \rho^4(\tau) + \dots \right]$$

Where

$$\rho^2(\tau) = \frac{1}{b_0^2} [g^2(\tau) + h^2(\tau)]$$

$$= J_0^2(\omega_m \tau)$$

Dropping terms beyond second degree, we get

$$R_x(\tau) \approx \frac{\pi}{2} b_0 \left[ 1 + \frac{1}{4} J_0^2(\omega_m \tau) \right] \quad (4.16)$$

The autocovariance function of  $r$  is defined for a stationary process as follows:

$$L_c(\tau) \equiv \langle [r(t) - \langle r \rangle][r(t+\tau) - \langle r \rangle] \rangle$$

$$= R_r(\tau) - \langle r \rangle^2 \quad (4.17)$$

In our case, considering the antenna gain factor  $G(\alpha)=1.5$ ,  $p(r)=(r/b_0)e^{-r^2/2b_0}$ . Thus,

$$\langle r \rangle = \frac{1}{b_0} \int_0^\infty r^2 e^{-\frac{r^2}{2b_0}} dr = \sqrt{\frac{\pi}{2}} b_0 \quad (4.18)$$

Substituting Equations 4.16 and 4.18 into Equation 4.17, we have

$$L_e(\tau) \approx \frac{\pi}{8} b_0 J_0^2(\omega_m \tau) \quad (4.19)$$

#### 4.3.4 Spatial Correlation of the Envelope

Let us consider the correlation of the envelope at two places with distance  $d$  under the same conditions assumed so far. Since we assume that the distribution of the arrival angle is uniform and the receiver is moving at a speed  $v$ , the correlation function in terms of the time difference  $\tau$  is equivalent to that in terms of the distance  $d=v\tau$ . Inserting the relations  $d=v\tau$  and  $\omega_m=2\pi v/\lambda$  into Equation 4.19, we have

$$L_e(d) \approx \frac{\pi}{8} b_0 J_0^2\left(\frac{2\pi d}{\lambda}\right)$$

The normalized autocovariance function becomes

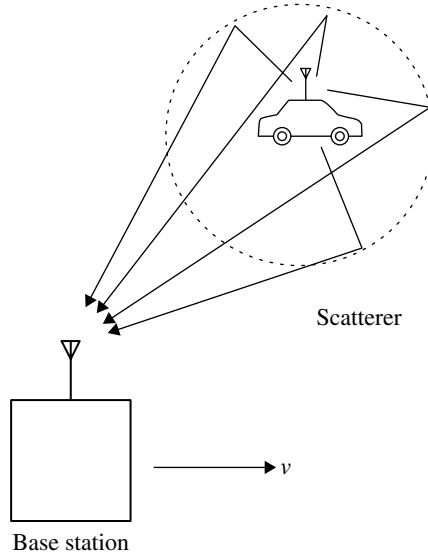
$$L_{en}(d) = \frac{L_e(d)}{L_e(0)} \approx J_0^2\left(\frac{2\pi d}{\lambda}\right)$$

This relation is useful for determining the correlation function for a dual space-diversity system at a receiver. The correlation becomes zero for the spacing  $d=0.38\lambda, 0.88\lambda, \dots$

The above results usually cannot be applied for a base station, since the uniform distribution of the arrival angle is hardly satisfied. Because of the height of the base station antenna, there are few scattering objects around the base station. The assumption that the receiver is moving can be equivalently satisfied by considering that the base station is moving instead of the mobile station. The spatial correlation at a base station is analyzed in Refs. [3] and [5] with a system model shown in Figure 4.5. Since the arrival angle is not uniformly distributed, the spread of the Doppler frequencies becomes small and the spatial correlation at a base station becomes high. A spatial separation on the order of ten times the wavelength is required for a base station diversity system.

#### 4.3.5 Random Frequency Modulation

The fast fading causes random variation in the phase  $\theta$  or its derivative  $\dot{\theta}$ , the instantaneous frequency as well as in the envelope  $r$ . The random variation of  $\dot{\theta}$  is equivalent to frequency modulation by the random signal.



**FIGURE 4.5** Analytical model for spatial correlation at a base station.

The joint density function is given as [3]:

$$p(r, \dot{r}, \theta, \dot{\theta}) = \frac{r^2}{4\pi^2 b_0 b_2} \exp \left[ -\frac{1}{2} \left( \frac{r^2}{b_0} + \frac{\dot{r}^2}{b_2} + \frac{r^2 \dot{\theta}^2}{b_2} \right) \right]$$

where the dot denotes the time derivative and  $b_1 = 0$  is assumed.  $b_n$  are the moments defined as

$$b_n = (2\pi)^n \int_{f_c - f_m}^{f_c + f_m} S(f) (f - f_c)^n df$$

For electric field,  $b_n$  is as follows:

$$b_n = \begin{cases} b_0 \omega_m^n \frac{1 \cdot 3 \cdot 5 \cdots (n-1)}{2 \cdot 4 \cdot 6 \cdots n} & (n = \text{even}) \\ 0 & (n = \text{odd}) \end{cases}$$

Integrating  $p(r, \dot{r}, \theta, \dot{\theta})$  over  $r$ ,  $\dot{r}$ , and  $\theta$ , we have

$$\begin{aligned} p(\dot{\theta}) &= \frac{1}{4\pi^2 b_0 b_2} \int_0^\infty dr \int_0^\infty d\dot{r} \int_0^{2\pi} d\theta r^2 \exp \left[ -\frac{1}{2} \left( \frac{r^2}{b_0} + \frac{\dot{r}^2}{b_2} + \frac{r^2 \dot{\theta}^2}{b_2} \right) \right] \\ &= \frac{1}{2} \sqrt{\frac{b_0}{b_2}} \left( 1 + \frac{b_0}{b_2} \dot{\theta}^2 \right)^{-3/2} \end{aligned}$$

For the electric field  $b_2 / b_0 = \omega_m^2 / 2$ , we have

$$p(\dot{\theta}) = \frac{1}{\omega_m \sqrt{2}} \left[ 1 + 2 \left( \frac{\dot{\theta}}{\omega_m} \right)^2 \right]^{-3/2} \quad (4.20)$$

#### 4.4 DELAY SPREAD AND FREQUENCY-SELECTIVE FADING

The results for fast fading, discussed so far, are derived from the assumption that there is no difference between the arrival times of the multipath signal (Eqs. 4.9 and 4.10). Actually, difference in the multipath delays exists. This fact causes no significant effect on signal transmission as long as we consider a signal with bandwidth less than the inverse of the magnitude of the difference in the delay times. Otherwise the effects cannot be neglected and the channel causes linear distortion of the received signal. The transfer function is such that the amplitude and the phase show dependence on frequencies over of the signal spectrum.

The multipath channel is characterized with impulse response

$$h(t) = \sum_{n=1}^{\infty} c_n \delta(t - \tau_n)$$

where  $c_n$  are complex coefficients that express the amplitude of the paths and  $\tau_n$  are the time delays. Since the phase of  $c_n$  is not correlated with  $\tau_n$ , the power, that is,  $|c_n|^2$  versus  $\tau_n$  is considered.

The power delay profile becomes different depending on the environment. In a mountainous area, we often encounter a large delay difference since echoes from far mountains reach the receiver with large delay times. In a city area, where buildings are not so tall, the delay difference becomes smaller. Many measurements for the delay profile have been carried out [3, 5].

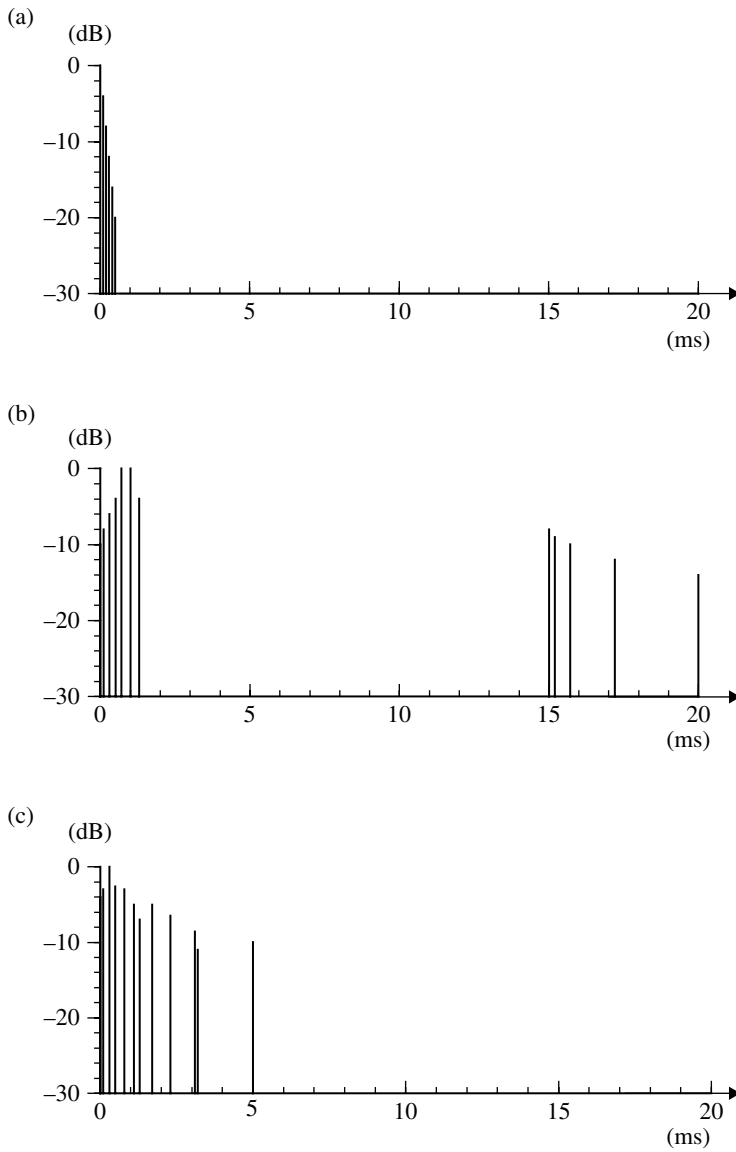
Some models for the delay profile are presented to evaluate signal transmission performance. Figure 4.6 shows the models used for the pan-European digital mobile telephone system. The two-path model is also often used for analytical treatments. The impulse response for this channel is

$$h(t) = c_1 \delta(t - \tau_1) + c_2 \delta(t - \tau_2) \quad (4.21)$$

where  $\tau_1$  and  $\tau_2$  are the delay times and the complex coefficients  $c_1$  and  $c_2$  are random variables whose amplitude obeys to the Rayleigh distribution and phase takes a uniform distribution. The power delay profile becomes

$$P(\tau) = |c_1|^2 \delta(\tau - \tau_1) + |c_2|^2 \delta(\tau - \tau_2) \quad (4.22)$$

Another assumption for the power delay profile is an exponential distribution. If we assume a continuous time delay, then



**FIGURE 4.6** Propagation model for the Pan-European digital cellular (GSM) system. (a) Typical rural area, (b) typical hilly terrain, and (c) typical urban area.

$$P(\tau) = \frac{1}{\sigma} e^{-\frac{\tau}{\sigma}} \quad (\tau \geq 0) \quad (4.23)$$

where  $\sigma$  is a coefficient denoting the degree of difference in time delay.

In order to represent the degree of difference in delay, the delay spread  $\Delta$  is defined as

$$\Delta \equiv \left[ \frac{\int_0^\infty (\tau - d_m)^2 P(\tau) d\tau}{\int_0^\infty P(\tau) d\tau} \right]^{1/2}$$

where  $d_m$  is an average delay defined as

$$d_m = \left[ \frac{\int_0^\infty \tau P(\tau) d\tau}{\int_0^\infty P(\tau) d\tau} \right] \quad (4.24)$$

The delay spread takes a value of around 3 ms for city area and up to 10 ms in hilly terrain. The delay spread can be calculated for the two-path model as:

$$\Delta = \left[ \frac{|c_1|^2 (\tau_1 - d_m)^2 + |c_2|^2 (\tau_2 - d_m)^2}{|c_1|^2 + |c_2|^2} \right]^{1/2}$$

where  $d_m = \frac{\tau_1 |c_1|^2 + \tau_2 |c_2|^2}{|c_1|^2 + |c_2|^2}$ .

When  $|c_1|^2 = |c_2|^2$ , then

$$\Delta = \frac{|\tau_1 - \tau_2|}{2}$$

For the exponential distribution, we have

$$\Delta = \sigma$$

#### 4.4.1 Coherence Bandwidth

Let us consider correlation between envelopes at the same time ( $\tau=0$  in Section 1.5.1 of Ref. [3]) and at different frequencies. The correlation coefficient is defined as

$$\rho(r_1, r_2) = \frac{\langle r_1 r_2 \rangle - \langle r_1 \rangle \langle r_2 \rangle}{\left\{ \left[ \langle r_1^2 \rangle - \langle r_1 \rangle^2 \right] \left[ \langle r_2^2 \rangle - \langle r_2 \rangle^2 \right] \right\}^{1/2}}$$

where  $r_1$  and  $r_2$  are the amplitudes of the signals at frequencies  $\omega_1$  and  $\omega_2$ , respectively. Introducing the time delay to Equations 4.9 and 4.10, the correlation function under the exponential power-delay profile of Equation 4.23 is given as [3]:



$$\rho(r_1, r_2) = \frac{1}{1 + s^2 \sigma^2}$$

where  $s = |\omega_1 - \omega_2|$  is the difference of the frequencies. The coherence bandwidth  $W_c$  is defined as the bandwidth where  $\rho(s) = 0.5$ . Thus,  $W_c = 1/2\pi\sigma$ .

For the two-path model, we normalize the power-delay profile (Eq. 4.22) as

$$p(\tau) = \frac{|c_1|^2 \delta(\tau - \tau_1) + |c_2|^2 \delta(\tau - \tau_2)}{|c_1|^2 + |c_2|^2}$$

Using Equation 4.24 and following the derivation process in Ref. [3] (calculating  $\lambda^2$  in Eqs. (1.5–17) and using  $\rho(r_1, r_2) \approx \lambda^2$  in Eqs. (1.5–26)), we have

$$\rho(r_1, r_2) \approx \frac{|c_1|^4 + |c_2|^4 + 2|c_1|^2 |c_2|^2 \cos[s(\tau_1 - \tau_2)]}{[|c_1|^2 + |c_2|^2]^2}$$

The coherence bandwidth depends on  $|c_1|$  and  $|c_2|$ . When  $|c_1| = |c_2|$ , it becomes

$$W_c = \frac{1}{4|\tau_1 - \tau_2|}$$

Measured coherence bandwidth in mobile radio channel is a few hundred kHz.

#### 4.4.2 Frequency-Selective Fading

If there is difference in time delay for a multipath channel, the channel has a transfer function that causes a (linear) distortion of the received signal. For example, let us consider the two-path model. Taking the Fourier transform of Equation 4.21, we have

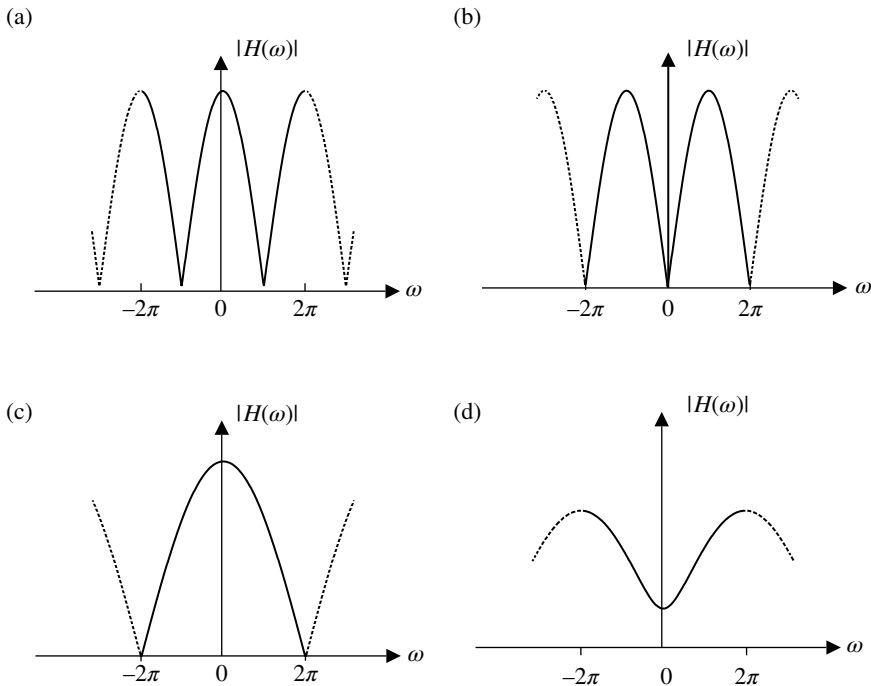
$$H(\omega) = c_1 e^{-j\omega\tau_1} + c_2 e^{-j\omega\tau_2}$$

As it is shown in Figure 4.7, the signal level fades at specific frequencies depending on the parameters  $c_1$ ,  $c_2$ , and  $\tau_1 - \tau_2$ . Here we assume that the coefficients  $c_1$  and  $c_2$  are given by

$$c_1, c_2 = T_c(t) + jT_s(t),$$

where  $T_c(t)$  and  $T_s(t)$  are defined by Equations 4.9 and 4.10, respectively. Then the channel becomes time varying and frequency selective. The varying speed is determined by the maximum Doppler frequency.

In the special case of  $\tau_1 = \tau_2$ , the channel becomes a Rayleigh fading channel with a flat transfer function.

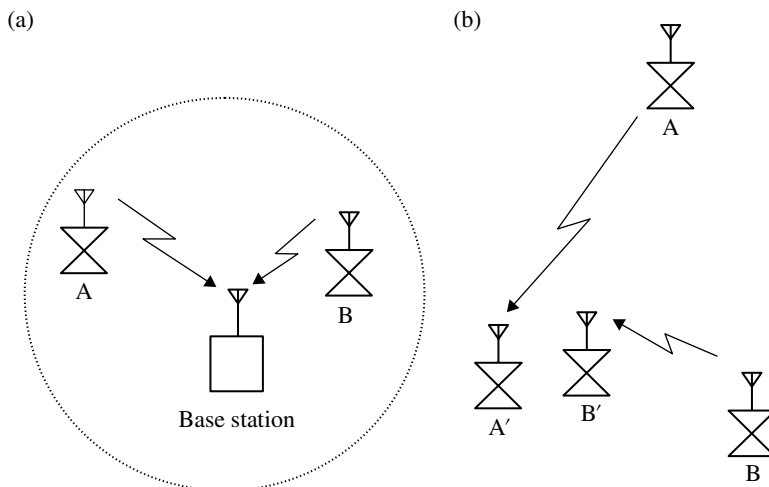


**FIGURE 4.7** Transfer functions of two-path frequency-selective fading channels. (a)  $C_1 = C_2 = 1$ ,  $\tau_1 - \tau_2 = 1$ , (b)  $C_1 = 1$ ,  $C_2 = -1$ ,  $\tau_1 - \tau_2 = 1$ , (c)  $C_1 = 1$ ,  $C_2 = 1$ ,  $\tau_1 - \tau_2 = 0.5$ , and (d)  $C_1 = 1$ ,  $C_2 = 0.5$ ,  $\tau_1 - \tau_2 = 0.5$ .

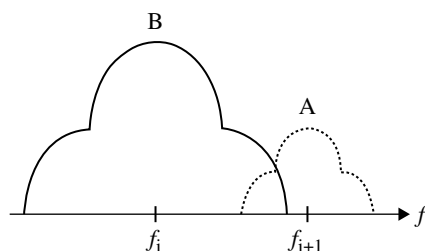
## 4.5 THE NEAR-FAR PROBLEM

The received signal level varies depending on the path length. The dynamic range of the received signal level can become as large as 70dB for path length of 100m to 10km; this value is found using Equation 4.3 with  $\alpha=3.5$ . The near-far problem stems from the wide dynamic range of the received level in mobile radio communication. Consider systems where two mobile stations are communicating with a base station or with other mobile stations as is illustrated in Figure 4.8a and b. At the base station, the signal levels received from the mobile stations A and B are quite different due to the difference in the path length. Let us assume that the mobile stations are using adjacent channels (Fig. 4.9). The out-of-band radiation of the signal from the mobile station B interferes with the signal from station A in the adjacent channel. The effect, called the adjacent channel interference, becomes serious when the difference in the received signal levels is high. For this reason, the out-of-band radiation must be kept small.

In a conventional mobile radio channel with channel separation of 25kHz, it is recommended that the relative interference power in the adjacent channel be less than -70dB [6]. The relative tolerable adjacent channel interference level can be



**FIGURE 4.8** Two situations in the near-far problem. (a) System with base station and (b) system without base station.



**FIGURE 4.9** Adjacent channel interference.

different depending on the systems. A system with automatic transmit power control, where the received signal level is controlled within a given range, can tolerate a high relative adjacent channel interference level.

The near-far problem becomes more important for code-division multiple access systems where spread spectrum signals are multiplexed on the same frequency using low cross-correlation codes (Section 6.4).

## 4.6 COCHANNEL INTERFERENCE

The key concept of cellular systems is spatial reuse of channels: the same channel is assigned to different cells, where the probability of cochannel interference between those cells is less than a given value. The probability of cochannel interference is defined as the probability that the desired signal level (envelope)  $r_d$  drops below a

value proportional to the interfering (or undesired) signal level (envelope)  $r_u$ , that is,  $P_c = \text{Prob}(r_d \leq \alpha r_u)$ , where  $\alpha$  is a protection ratio.

Let us assume that the desired and the interfering signals are independent of each other. We denote the probability density functions for the desired and the interfering signals as  $p_1(r_1)$  and  $p_2(r_2)$ , respectively. Then  $P_c$  is given as

$$\begin{aligned} P_c &= \int_0^{\infty} \text{Prob}(r_1 = x) \text{Prob}\left(r_2 \geq \frac{x}{\alpha}\right) dx \\ &= \int_0^{\infty} p_1(r_1) \int_{r_1/\alpha}^{\infty} p_2(r_2) dr_2 dr_1 \end{aligned}$$

In the following, the probability of cochannel interference is calculated for Rayleigh fading, shadowing, and the combined Rayleigh fading and shadowing signals.

#### 4.6.1 Rayleigh Fading

The probability distribution function of the amplitude of a signal subjected to the Rayleigh fading is given by Equation 4.11a. In this case, the probability of the cochannel interference can be calculated as

$$\begin{aligned} P_{\text{CR}} &= \int_0^{\infty} \frac{r_1}{b_1} e^{-\frac{r_1^2}{2b_1}} \int_{r_1/\alpha}^{\infty} \frac{r_2}{b_2} e^{-\frac{r_2^2}{2b_2}} dr_2 dr_1 \\ &= \frac{1}{1 + \alpha^{-2} b_1/b_2} \end{aligned}$$

where  $b_1$  and  $b_2$  are the average power of the desired and undesired signals, respectively, and  $\alpha$  is the protection ratio.

#### 4.6.2 Shadowing

The probability distribution function of shadowing is given by Equation 4.7. Then we can calculate the probability of the cochannel interference  $P_{\text{cs}}$  (see Appendix 4.B).

$$\begin{aligned} P_{\text{cs}} &= \text{Prob}(x_1 \leq x_2 + \beta) \\ &= \frac{1}{\sqrt{\pi}} \int_b^{\infty} e^{-u^2} du = \frac{1}{2} \text{erfc}(b) \end{aligned} \tag{4.25}$$

where  $x_1$  and  $x_2$  are the desired and undesired signal envelopes in decibels, respectively, and  $\beta$  is the protection ratio in decibels. In Equation 4.25,  $b$  is given as

$$b = \frac{x_{1m} - x_{2m} - \beta}{2\sigma}$$

where  $x_{1m} = \langle x_1 \rangle$ ,  $x_{2m} = \langle x_2 \rangle$ , and  $\sigma$  is the standard deviation in decibels.

### 4.6.3 Combined Fading and Shadowing

The probability density function of the signal envelope  $r$  under the combined fading and shadowing is given with Equations 4.7 and 4.11a as

$$p_i(r) = \int_{-\infty}^{\infty} \frac{r}{b} e^{-\frac{r^2}{2b}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-x_{im})^2}{2\sigma^2}} dx \quad (i=1,2) \quad (4.26)$$

where  $b$  is a function of  $x$ , as  $b = 10^{\frac{x}{10}}$ .

The interference probability under this condition is given as

$$\begin{aligned} P_{\text{CF\&S}} &= \int_0^{\infty} \text{Prob}(r_1 = y) \text{Prob}\left(r_2 \geq \frac{y}{\gamma}\right) dy \\ &= \int_0^{\infty} p_1(r_1) \int_{r_1/\gamma}^{\infty} p_2(r_2) dr_2 dr_1 \end{aligned} \quad (4.27)$$

where  $r_1$  and  $r_2$  are the desired and undesired signal envelopes and  $\gamma$  is the protection ratio. Equation 4.27 can be reduced to (see Appendix 4.C)

$$P_{\text{CF\&S}} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{1}{1 + 10^{(x_{1m} - x_{2m} - R + 2\sigma u)/10}} e^{-u^2} du \quad (4.28)$$

where  $u$  is an internal variable and  $R = 20 \log \gamma$ .

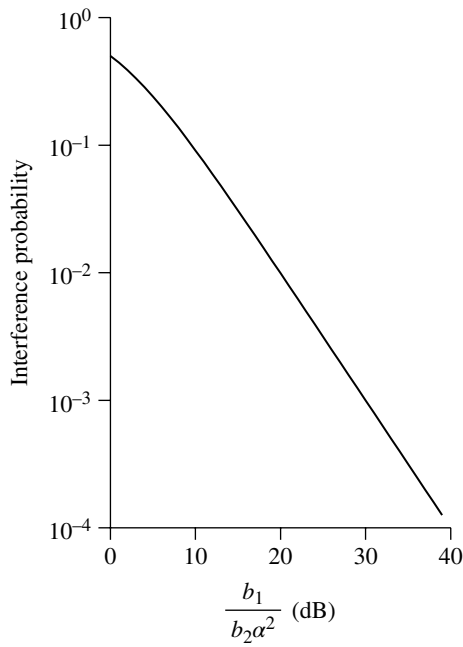
### 4.6.4 Discussion

The cochannel interference probabilities are shown in Figures 4.10, 4.11, and 4.12. Typically, for a fast fading environment, the cochannel interference under shadowing (Eq. 4.25) is used as a measure of system quality, for example, bit error rate. For a slow fading environment, the cochannel interference under both fading and shadowing are taken into consideration.

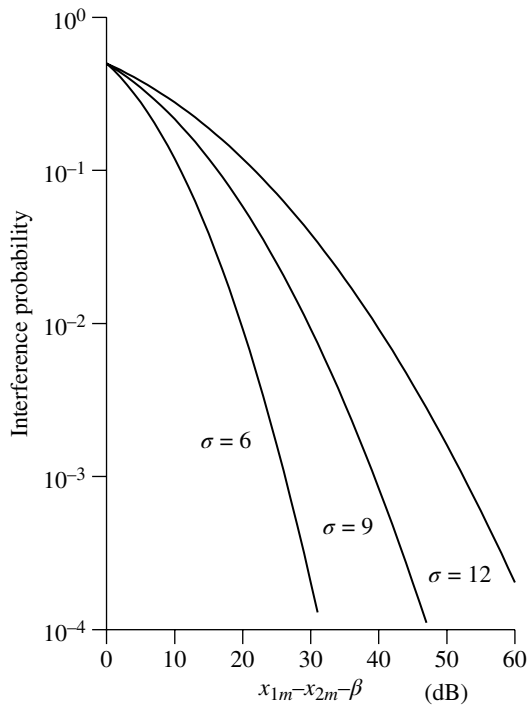
The above results for the cochannel interference probability can be applied to the adjacent channel interference by using appropriate values for the protection ratio.

## 4.7 RECEIVE POWER DISTRIBUTION AND RADIO CHANNEL DESIGN

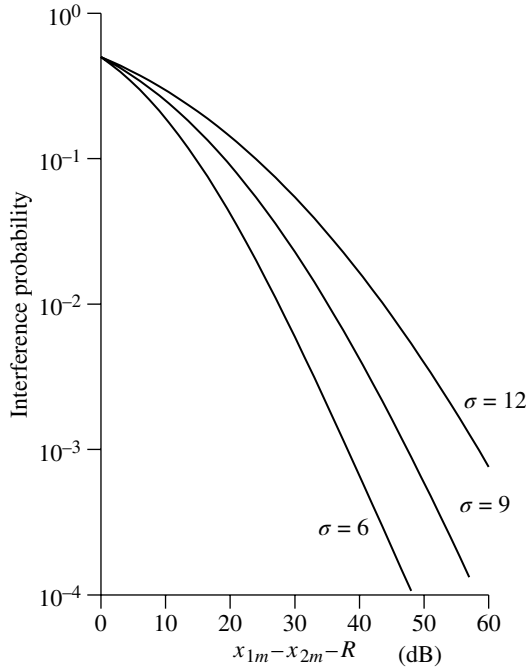
Receiver performance is determined if receive signal power and noise power are given. In an out-of-sight radio channel, the receiver performance is expected only statistically due to fading phenomena in the channel. Here we give cumulative distribution of receive signal power in fading channels and a radio link channel design example.



**FIGURE 4.10** Cochannel interference probability under Rayleigh fading.



**FIGURE 4.11** Cochannel interference probability under log-normal shadowing.



**FIGURE 4.12** Cochannel interference probability under fading and shadowing.

#### 4.7.1 Receive Power Distribution

We take into consideration the Rayleigh fading, shadow fading, path-loss, and a uniform distribution of mobile stations in a base station coverage area. Then the probability distribution function of receive power  $q$  is expressed as

$$p(q) = p_R(q, b) p_s(x, x_0) p_d(l)$$

where  $p_R(q, b)$  and  $p_s(x, x_0)$  are probability density functions in the Rayleigh fading and the shadow fading given with Equations 4.11b and 4.7, respectively. Where  $b = \langle q \rangle = 10^{x/10}$  and  $x_0 = \langle x \rangle$ .  $x_0$  becomes a function of distance  $l$  between a base station and a mobile station. If we assume a uniform distribution of mobile stations in an area ranging from  $R_0$  to  $R$ , the probability density function of the communication distance  $l$  is given as

$$p_d(l) = \frac{2l}{R^2 - R_0^2} \quad (4.29)$$

The cumulative distribution of receive power  $q$  is expressed as

$$\begin{aligned} P(q_0) &= \text{Prob}(q \leq q_0) \\ &= \int_0^{q_0} p_R(q, b) dq \int_{-\infty}^{\infty} p_s(x, x_0) dx \int_{R_0}^R p_d(l) dl \end{aligned}$$

Integrating over the Rayleigh fading, we have

$$\begin{aligned} P_R(q_0) &= \int_0^{q_0} p_R(q, b) dq \\ &= 1 - e^{-q_0/b} \end{aligned}$$

When  $M$ -branch maximal ratio combining (MRC) diversity (Section 7.1.1) is used, integrating Equation 7.3, we get

$$P_R(q_0) = 1 - e^{-q_0/b} \sum_{k=1}^M \frac{(q_0/b)^{k-1}}{(k-1)!} \quad (4.30)$$

Thus, we have

$$P(q_0) = \int_{-\infty}^{\infty} P_R(q_0, x) p_s(x, x_0) dx \int_{R_0}^R p_d(l) dl$$

The cumulative power distribution in dB for the Rayleigh fading is shown in Figure 4.13a. The power is normalized with the average power  $b = \langle q \rangle$ . The normalized power drops below a threshold value of  $-13$  dB with the cumulative probability of 5% (outage probability). When we use the two-branch MRC diversity, the threshold value of the normalized power increases to  $-4.5$  dB for the same outage probability of 5%.

Figure 4.13b shows the cumulative power distribution when the Rayleigh and shadow fading (with a variance of 8 dB) are considered simultaneously. The power is normalized with the average power  $10^{x_0/10}$ . Since the shadow fading is combined with the Rayleigh fading, the threshold normalized power for outage probability of 5% becomes lower as  $-19$  dB. The diversity improves the threshold value up to  $-12.5$  dB.

Figure 4.13c shows the results when the distribution of mobile stations is furthermore considered, where the path loss is expressed in dB as  $\alpha + \beta \log(l)$  ( $l$  in km,  $\beta = 33$ ). The power is normalized with average power  $P_0$  at  $l = 1$  km, where

$$P_0 = P_t G_b G_m 10^{-\alpha/10}$$

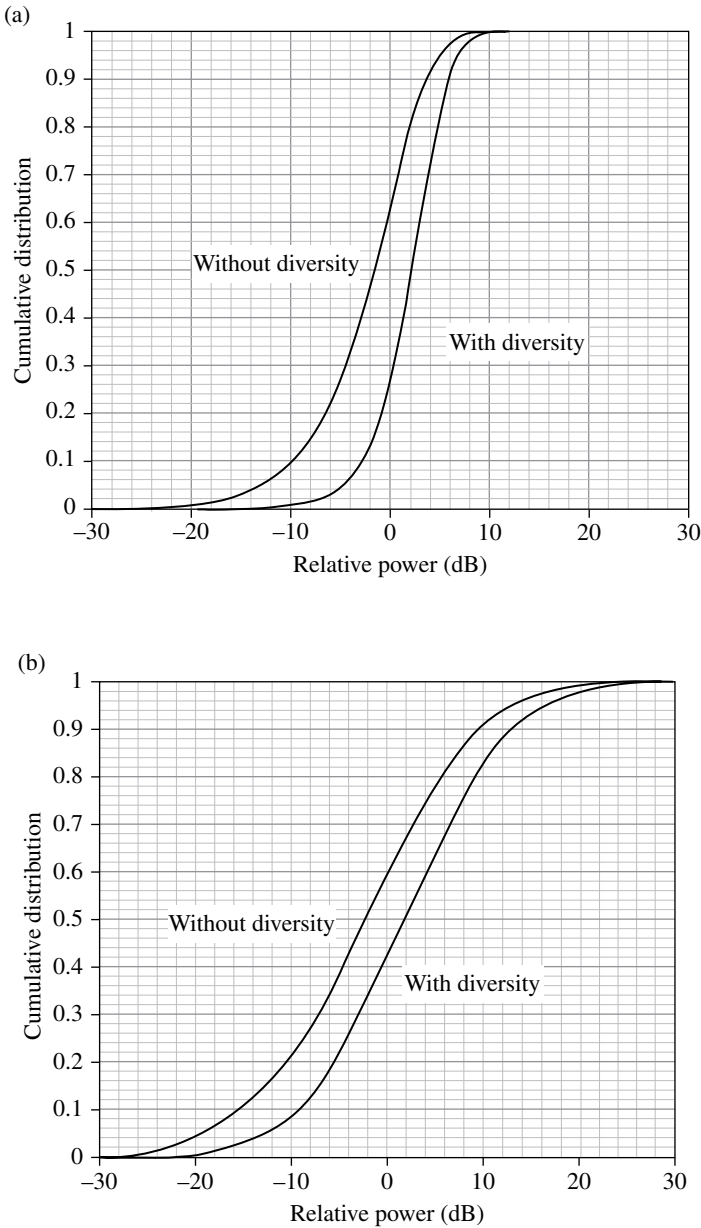
where  $P_t$  is the transmit power, and  $G_b$  and  $G_m$  are the antenna gains for base station and mobile stations, respectively.

#### 4.7.2 Channel Link Design

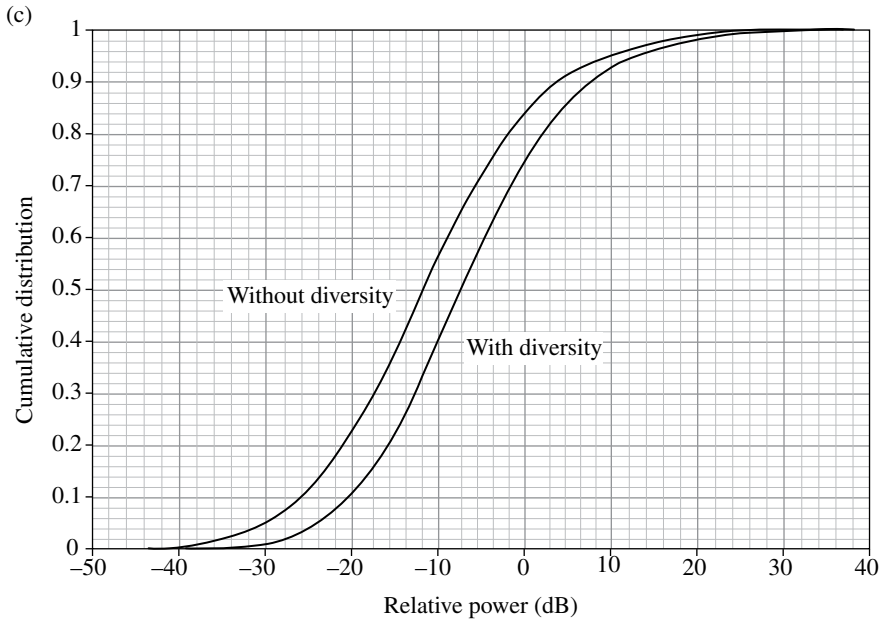
Average receive power given with Equation 4.1 is rewritten as

$$P_r = P_t G_b G_m L_c$$





**FIGURE 4.13** Cumulative distributions of receive power. (a) Under Rayleigh fading, (b) under Rayleigh and shadow fading, and (c) (*Continued*)



**FIGURE 4.13** Cumulative distributions of receive power (c) under Rayleigh, shadowing, and pass loss ( $f_c = 800$  MHz, distance = 3 km).

where  $L_c$  denotes path loss. We assume that the pass loss in decibel is approximated with Equation 4.4 as

$$\begin{aligned} [L'_c] &\equiv 10 \log \left( \frac{1}{L_c} \right) \\ &= \alpha + \beta \log(l) \quad (l \text{ in km}) \end{aligned}$$

where constants  $\alpha$  and  $\beta$  are determined if carrier frequency  $f_c$ , base station height  $h_b$ , and mobile station height  $h_m$  are given (Eqs. 4.4–4.6).

Noise power  $N$  at the input of a receiver is the summation of channel noise  $N_c$  and input equivalent noise power  $N_e$  generated in the receiver ( $N = N_c + N_e$ ). Following from the arguments in Section 2.2.1,  $N_e$  is given as

$$\begin{aligned} N_e &= kT_e B \\ &= k(NF - 1)T_a B \end{aligned}$$

where  $k$  ( $= 1.38 \times 10^{-23}$  W/Hz/K) is the Boltzmann constant,  $T_e$  the input equivalent noise temperature,  $B$  the equivalent noise bandwidth, NF the noise figure, and  $T_a = 290^\circ\text{K}$ . Denoting equivalent temperature of the channel noise by  $T_c$ , we have  $N_c = kT_c B$ . Thus, we get

$$N = (NF - 1 + T_c/T_a) kT_a B$$

If  $T_c = T_a$ , we have  $N = NF kT_a B$ .

Using the dB notation, we express the signal power  $S (=Pr)$  in dBm as

$$\begin{aligned} [S] &\equiv 10 \log S \\ &= 10 \log (P_t G_b G_m L_c) \\ &= [P_t] + [G_b] + [G_m] + [L_c] \\ &= [P_t] + [G_b] + [G_m] - \alpha - \beta \log(l) \quad (\text{dBm}) \end{aligned}$$

and the noise power as

$$\begin{aligned} [N] &= 10 \log ((NF - 1 + T_c/T_a) kT_a B) \\ &= -144 + 10 \log B(\text{kHz}) + 10 \log (NF - 1 + T_c/T_a) \quad (\text{dBm}) \end{aligned}$$

where the noise bandwidth is in kHz.

The signal to noise power ratio in dB is given as

$$\left[ \frac{S}{N} \right] = [P_t] + [G_b] + [G_m] - \alpha - \beta \log(l) + 144 - 10 \log B(\text{kHz}) - 10 \log (NF - 1 + T_c/T_a)$$

*Example:* We assume system parameters as shown in Table 4.1. From Equation 4.4, path loss becomes  $[L_c] = 120 + 33 \log(l)$  (dB). We investigate communication performance when a mobile station is located at a distance of 2 km from a base station. Then signal to noise power ratios  $[S/N]$  becomes 34 dB in the downlink and 25 dB in the uplink. When a communication distance is given, relative power distribution is shown in Figure 4.13b. The reference value (0 dB) of the distribution corresponds to the above  $[S/N]$  values.

The probability that  $S/N$  values drop below a threshold value, which is a required minimum value to satisfy a given communication quality, is called the outage probability. The required  $S/N$  depends on communication systems, especially on modulation system. If we take an outage probability of 5%, then the deterioration of  $S/N$  due to the Rayleigh and shadow fading becomes 19 and 13 dB without and with the diversity, respectively. If the diversity is employed only at the base station, the threshold  $S/N$  value for 5% outage probability becomes 12 dB in the uplink and 15 dB in the

**TABLE 4.1 Assumed System Parameters**

Area	Urban, large city
Carrier frequency: $f_c$	800 MHz
Antenna heights	70 m (base), 1.5 m (mobile)
Transmit power: $P_t$	10 W (base), 0.5 W (mobile)
Antenna gains: $G_b, G_m$	17 dB (base), 0 dB (mobile)
Noise figures: $NF$	3 dB (base), 7 dB (mobile)
External noise temperature: $T_c$	290°K
Equivalent noise bandwidth: $B$	1 MHz
Shadow fading variance: $\sigma$	8 dB

downlink. Then the downlink can support two times wider signal bandwidth than the uplink, if the required  $S/N$  is the same for the up- and downlink.

#### APPENDIX 4.A PROPAGATION LOSS FORMULA

The free space propagation relation known as the Friss's formula is given as

$$P_r = \left( \frac{\lambda}{4\pi d} \right)^2 G_r G_t P_t$$

which is usually derived by assuming an equivalent (vector) length for a microscopic dipole antenna and an arbitrary linear antenna. Here, we present another method which is more intuitively understandable. We use the Kirchhoff · Huygens approximation as

$$E_r \approx \frac{jk}{4\pi} \frac{e^{-jkr}}{r} (1 + \cos\theta) \int_S E_t e^{jk\rho \cdot r} dS$$

where  $E_r$  and  $E_t$  are the electric fields at receiving point and at transmitting points, respectively.  $S$  represents transmitting space region,  $\rho$  denotes a position vector in the transmitting region,  $r$  the vector that connects the transmitting point and receiving point,  $r$  its distance,  $\theta$  the angle between  $\rho$  and  $r$  measured from the normal direction, and  $k$  the propagation constant. The electromagnetic field at receiving antenna region is approximated as a plane wave if the transmit and receiving antenna distance is long enough, in spite of transmitting antenna type. Now we assume an imaginary transmit antenna that produces a plane wave of which electric field takes a value of  $E_t$  in a given area of size  $\sigma_t$  at transmit antenna region and zero outside the area. This kind of antenna is well represented by a big aperture antenna. (When transmit power is  $P_t$ , then the received power at that antenna for a plane wave whose electric field strength is  $E_t$  becomes  $P_r$ , due to the reciprocity of radio wave.)

The strength of electric field which is transmitted in a direction perpendicular to the transmit antenna aperture ( $\theta = 0$  and therefore  $\rho \cdot r = 0$ ) is given at the distance  $r = d$  by the Kirchhoff · Huygens formula as

$$E_r \approx \frac{jk}{2\pi} \frac{e^{-jkd}}{d} E_t \sigma_t, \text{ and therefore } |E_r| = \frac{\sigma_t}{\lambda d} |E_t| \quad (4.A.1)$$

where  $\lambda = 2\pi/k$  is the wave length.

Power density  $P_{t0}$  (W/m<sup>2</sup>) at the transmit antenna aperture  $\sigma_t$  is given as

$$P_{t0} = \frac{|E_t|^2}{Z_0} \quad (4.A.2)$$

where  $Z_0$  denotes wave impedance. The transmit power is expressed as

$$P_t = \sigma_t P_{t0} \quad (4.A.3)$$

From the definition of antenna gain and the law of the conservation of power, power density  $P_{r0}$  at the receiving antenna is given as

$$\frac{|E_r|^2}{Z_0} = \frac{P_t G_t}{4\pi d^2} \quad (4.A.4)$$

From Equations 4.A.1–4.A.4, we get

$$\sigma_t = \frac{\lambda^2}{4\pi} G_t \quad (4.A.5)$$

We define the equivalent receive cross section of a receive antenna as follows:

$$\sigma_r = \frac{P_r}{P_{r0}} \quad (4.A.6)$$

where  $P_{r0}$  denotes the receive power density and is expressed as follows:

$$P_{r0} = \frac{|E_r|^2}{Z_0} \quad (4.A.7)$$

From Equations 4.A.4, 4.A.6, and 4.A.7, we get the following relation:

$$P_r = \sigma_r P_{r0} = \frac{\sigma_r G_t P_t}{4\pi d^2} \quad (4.A.8)$$

By interchanging the transmit and receive antennas in the above argument and from the reciprocity of an antenna system, we get the following relation:

$$P_{12} = \frac{\sigma_2 G_1 P_0}{4\pi d^2} = P_{21} = \frac{\sigma_1 G_2 P_0}{4\pi d^2} \quad (4.A.9)$$

From the above equation, we get:

$$\sigma_2 G_1 = \sigma_1 G_2$$

With this equation and Equation 4.A.5, we have:

$$\sigma_r = \frac{\lambda^2}{4\pi} G_r \quad (4.A.10)$$

Inserting the above equation into Equation 4.A.8, we get  $P_r = \left(\frac{\lambda}{4\pi d}\right)^2 G_r G_t P_t$ .

The equivalent transmit antenna (cross section) is not a real antenna (cross section) and is based on an imaginary concept. And therefore the above result is valid for any antenna. This concept is similar to the equivalent antenna (vector) length of a linear antenna.

Equation 4.A.8 means that received power never depends on the wavelength. Rewriting Equation 4.A.10, we have  $G_r = 4\pi \frac{\sigma_r}{\lambda^2}$ . The antenna gain is proportional to the equivalent cross section and inversely proportional to the square value of wavelength. This means the fact that if we have the same equivalent cross section for the same received power, the antenna gain becomes high with increase in frequency: a high gain antenna is difficult to use in some circumstance, especially in mobile communications.

## APPENDIX 4.B INTERFERENCE PROBABILITY UNDER SHADOWING

Denote the desired and undesired signal levels in decibels as  $x_1$  and  $x_2$ . Then the probability density function of  $x_1$  and  $x_2$  are given from Equation 4.7 as

$$\begin{aligned} p_1(x_1) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1 - x_{1m})^2}{2\sigma^2}} \\ p_2(x_2) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_2 - x_{2m})^2}{2\sigma^2}} \quad (-\infty < x_1, x_2 < \infty) \end{aligned} \quad (4.B.1)$$

We define a new variable  $z = x_1 - x_2$ . Then, we have

$$P_{cs} = \text{Prob}(z \leq \beta) \quad (4.B.2)$$

The probability density function of  $z$  is given by

$$p(z) = \int_{-\infty}^{\infty} p_1(z + x_2) p_2(x_2) dx_2 \quad (-\infty < z < \infty) \quad (4.B.3)$$

Using (4.B.1), we have

$$\begin{aligned} p(z) &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} e^{-\frac{(z+x_2 - x_{1m})^2}{2\sigma^2} - \frac{(x_2 - x_{2m})^2}{2\sigma^2}} dx_2 \\ &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} e^{-\frac{\left(x_2 + \frac{z - x_{1m} - x_{2m}}{2}\right)^2}{\sigma^2}} e^{-\frac{(z - x_{1m} + x_{2m})^2}{4\sigma^2}} dx_2 \end{aligned}$$

Performing the integration over  $x_2$ , and using a formula,  $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$ , we have

$$p(z) = \frac{1}{2\sqrt{\pi}\sigma} e^{-\frac{(z - x_{1m} + x_{2m})^2}{4\sigma^2}}$$

Then, we get

$$P_{cs} = \frac{1}{2\sqrt{\pi}\sigma} \int_{-\infty}^{\beta} e^{-\frac{(z - x_{1m} + x_{2m})^2}{4\sigma^2}} dz = \frac{1}{\sqrt{\pi}} \int_{(x_{1m} - x_{2m} - \beta)/2\sigma}^{\infty} e^{-u^2} du$$

## APPENDIX 4.C INTERFERENCE PROBABILITY UNDER COMBINED FADING AND SHADOWING

With Equations 4.27 and 4.28, we have

$$P_{\text{CF\&S}} = \int_0^\infty \int_{-\infty}^\infty \frac{r_1}{b_1} e^{-\frac{r_1^2}{2b_1}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1-x_{1m})^2}{2\sigma^2}} dx_1 \\ \times \int_{r_1/\gamma}^\infty \int_{-\infty}^\infty \frac{r_2}{b_2} e^{-\frac{r_2^2}{2b_2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_2-x_{2m})^2}{2\sigma^2}} dx_2 dr_2 dr_1$$

The integration over  $r_2$  is readily calculated and we have

$$P_{\text{CF\&S}} = \int_{-\infty}^\infty \int_{-\infty}^\infty \frac{r_1}{b_1} e^{-\frac{b_1\gamma^{-2}+b_2}{b_1b_2} \frac{r_1^2}{2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1-x_{1m})^2}{2\sigma^2}} dx_1 \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_2-x_{2m})^2}{2\sigma^2}} dx_2 dr_1$$

Performing the integration over  $r_1$ , we get

$$P_{\text{CF\&S}} = \frac{1}{2\pi\sigma^2} \int_{-\infty}^\infty \int_{-\infty}^\infty \frac{1}{1+\gamma^{-2}b_1/b_2} e^{-\frac{(x_1-x_{1m})^2 + (x_2-x_{2m})^2}{2\sigma^2}} dx_1 dx_2 \quad (4.C.1)$$

Using new variables  $y_1 = x_1 - x_{1m}$  and  $y_2 = x_2 - x_{2m}$  and expressing  $b_1 = 10^{x_1/10}$  and  $b_2 = 10^{x_2/10}$ , Equation 4.C.1 can be rewritten as

$$P_{\text{CF\&S}} = \frac{1}{2\pi\sigma^2} \int_{-\infty}^\infty \int_{-\infty}^\infty \frac{1}{1+10^{(y_1-y_2+x_{1m}-x_{2m}-R)/10}} e^{-\frac{(y_1+y_2)^2 + (y_1-y_2)^2}{4\sigma^2}} dy_1 dy_2$$

Using new variables  $y_1 + y_2 = \sqrt{2} t_1$  and  $y_1 - y_2 = \sqrt{2} t_2$  and integrating over  $t_1$ , we have a single integral as

$$P_{\text{CF\&S}} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty \frac{1}{1+10^{(x_{1m}-x_{2m}-R+\sqrt{2}t_2)/10}} e^{-\frac{t_2^2}{2\sigma^2}} dt_2$$

Letting  $u = t_2/\sqrt{2}\sigma$ , we have Equation 4.28.

## REFERENCES

- [1] Okumura Y, Ohmori E, Kawano T, Fukua K. Field strength and its variability in UHF and VHF land-mobile radio service. Rev Electr Commun Lab 1968;16.
- [2] Hata M. Empirical formula for propagation loss in land mobile radio services. IEEE Trans Veh Technol 1980;VT-29:317-325.

- [3] Jakes WC, editor. *Microwave Mobile Communications*. New York: John Wiley & Sons, Inc.; 1974.
- [4] Mitchel RL. Performance of the log-normal distribution. *J Opt Soc Am* 1968;58: 1267–1272.
- [5] Lee WCY. *Mobile Communications Engineering*. New York: McGraw-Hill; 1982.
- [6] CCIR Recommendation 478–2. (CCIR: Comité Consultative International des Radiocommunication).



---

# 5

---

## ELEMENTS OF DIGITAL MODULATION

Modulation is a process where a baseband signal is converted into a radio frequency (RF) signal. Demodulation is an inverse process, where the baseband signal is recovered from the RF signal. An RF signal has two degrees of freedom, namely, amplitude and phase, or in-phase and quadrature components. Modulation is carried out by varying these components according to the baseband signal. Digital modulation is different from analog in the sense that the digital signal takes finite discrete levels. Due to this feature, some particular techniques can be applied to the modulation and demodulation in digital communications.

The requirements for digital modulation are the following: narrow bandwidth of the modulated signal, low bit error rate, and easy implementation of the modulation/demodulation circuits.

### 5.1 DIGITALLY MODULATED SIGNALS

This section describes briefly some general topics of digital modulation and demodulation. The reader may refer many books [1–8] for further details.

A modulated signal can be expressed generally as

$$\begin{aligned} f(t) &= A(t) \cos \{ \omega_c t + \varphi(t) \} \\ &= \operatorname{Re} \left\{ A(t) e^{j[\omega_c t + \varphi(t)]} \right\} = \operatorname{Re} \left\{ f_b(t) e^{j\omega_c t} \right\} \end{aligned}$$

where  $A(t)$  is the amplitude of the envelope,  $\varphi(t)$  the phase,  $\omega_c$  the (angular) frequency of the carrier signal, and  $f_b(t) = A(t)e^{j\varphi(t)}$ , the complex baseband signal. We can rewrite the above equation as

$$f(t) = x(t)\cos(\omega_c t) - y(t)\sin(\omega_c t)$$

where  $x(t) = A(t)\cos\varphi(t)$  and  $y(t) = A(t)\sin\varphi(t)$  are the in-phase and quadrature components, respectively. The carrier signal has two freedoms, namely  $A(t)$  and  $\varphi(t)$  or  $x(t)$  and  $y(t)$ . The time differentiation of  $\varphi(t)$  is the instantaneous (angular) frequency.

From the above expression, one cannot distinguish a digitally modulated signal from an analog modulated signal. The difference between the two appears only in detailed expressions of  $A(t)$  and  $\varphi(t)$  or  $x(t)$  and  $y(t)$ . Modulation techniques amplitude modulation (AM), phase modulation (PM), and frequency modulation (FM) (Section 2.1.5) in analog systems are called amplitude shift keying (ASK), phase shift keying (PSK), frequency shift keying (FSK) in digital systems, respectively. The quadrature amplitude modulation (QAM), where the modulating signals  $x(t)$  and  $y(t)$  are independent from each other, is used widely for digital modulation.

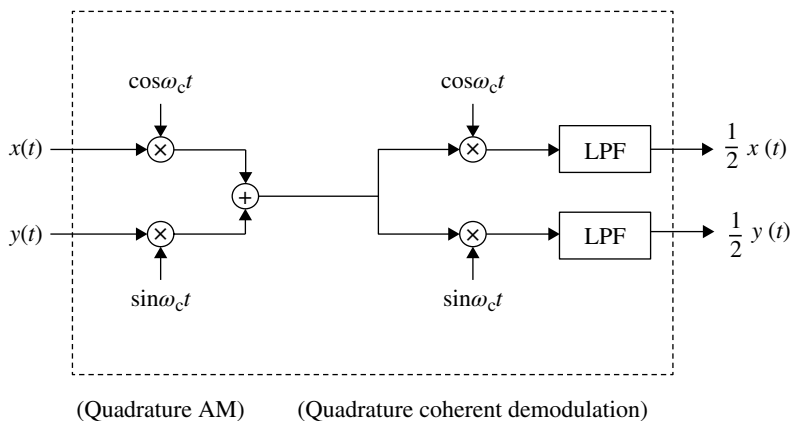
## 5.2 LINEAR MODULATION VERSUS CONSTANT ENVELOPE MODULATION

ASK and QAM fall into the category of linear modulation. Linear modulation by definition can be generated by multiplying the baseband modulating signal(s) by a carrier signal. Thus, the baseband spectrum of a linear modulated signal is symmetrically placed at the upper and lower sides of the carrier frequency. However, PSK with constant amplitude and FSK can be called nonlinear or constant envelope modulation techniques. In this type of system, the modulating process includes a nonlinear operation, viz., trigonometric function of the modulating signal. The spectrum of the nonlinear modulation signal is quite different from the baseband signal spectrum.

The phase of a digitally modulated PSK signal is selected from a set of phases corresponding to the input digital signal. It can be either a linear or a constant envelope (nonlinear) modulation. Most of the actual PSK signals are linearly modulated.

Linear modulation requires a linear channel. For example, a linear amplifier must be used for a transmit amplifier, otherwise the spectrum of the modulated signal spreads and the amplitude and/or phase component is distorted from nonlinearity of the circuit. In contrast to this, the constant envelope signal never requires a linear channel; for example, a saturated power amplifier can be used. Although the waveform of the modulated signal becomes distorted at the output of the power amplifier, it is recovered by passing through a bandpass filter. This filter removes the higher order frequency components generated in the saturated amplifier. This discussion is assessed in Appendix 5.A.

A modulated signal can be represented on a two-dimensional plane where the in-phase and quadrature components correspond to the horizontal and vertical axes. An expression on this plane is equivalent to the expression on the complex amplitude plane.



**FIGURE 5.1** Quadrature AM and its coherent demodulator.

The quadrature AM has independent in-phase and quadrature components and shows a spectrum efficiency two times higher than ordinary AM if the two components has the same spectrum bandwidth. This is owing to the fact that the two modulated signals corresponding to each component has their spectra overlapped with each other. In spite of the overlapping of the modulated signal spectra, they can be separated at the quadrature coherent demodulator output as in-phase and quadrature components (Fig. 5.1). The combined modulator and demodulator as a whole can be a time-invariant linear system.

## 5.3 DIGITAL MODULATIONS

### 5.3.1 Phase Shift Keying

A PSK signal takes one of the fixed phases corresponding to the digitally modulating signal. Phase constellations of binary PSK (BPSK), quaternary PSK (QPSK), and 8-level PSK (8PSK) are shown in Figure 5.2. A symbol of BPSK, QPSK, and 8PSK represents 1, 2, and 3 bits of digital signal, respectively.

The fact that the signal phase of a PSK signal takes specific values makes it possible to generate a carrier signal by frequency multiplication as is shown later. From the same fact, differential or delay detection (where the relative phase change between successive symbol times is detected) can be applied to a PSK signal.

#### 5.3.1.1 BPSK The BPSK signal is expressed as

$$s(t) = \sum_{n=-\infty}^{\infty} a_n h(t - nT) \cos \omega_c t \quad (a_n = \pm A)$$

where  $a_n$  takes  $A$  or  $-A$  (volt) corresponding to the digital signal,  $h(t)$  is the impulse response of the baseband filter, and  $\omega_c$  is the carrier frequency. The block diagram of a BPSK modulator is shown in Figure 5.3.

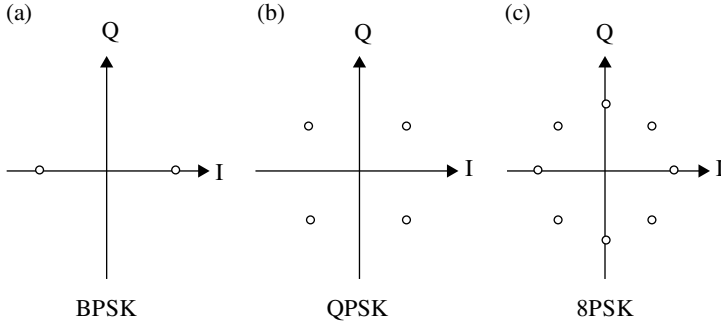


FIGURE 5.2 Signal constellation.

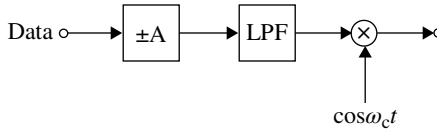


FIGURE 5.3 Block diagram of BPSK modulator.

**5.3.1.2 QPSK** The QPSK signal is described as

$$\begin{aligned}
 s(t) &= \sum_{n=-\infty}^{\infty} a_n h(t - nT_s) \cos \omega_c t + \sum_{n=-\infty}^{\infty} b_n h(t - nT_s) \sin \omega_c t \\
 &= \operatorname{Re} \left\{ \sum_{n=-\infty}^{\infty} (a_n - jb_n) h(t - nT_s) e^{j\omega_c t} \right\} \quad (a_n, b_n = \pm A)
 \end{aligned} \tag{5.1}$$

where  $T_s$  is the symbol duration and  $a_n, b_n$  are the input data. Figure 5.4 shows the block diagram of a QPSK modulator.

**5.3.1.3  $\pi/2$  Shifted BPSK** The signal phase constellation is shown in Figure 5.5. The phase is selected such that the signal never has the same phase for two successive symbols. The block diagram of the modulation is shown in Figure 5.6. The  $\pi/2$  shifted BPSK signal is expressed as

$$\begin{aligned}
 s(t) &= \sum_{n=-\infty}^{\infty} a_{2n} h(t - 2nT) \cos \omega_c t \\
 &\quad + \sum_{n=-\infty}^{\infty} a_{2n+1} h[t - (2n+1)T] \sin \omega_c t \quad (a_n = \pm A)
 \end{aligned} \tag{5.2}$$

where  $T$  is the bit duration. It must be noted that the bandwidth of the low-pass filter whose impulse response is  $h(t)$  is the same as that of the BPSK. Thus, the power spectra are the same for  $\pi/2$  shifted BPSK and BPSK.

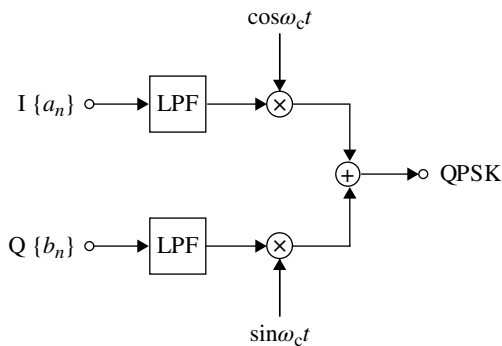
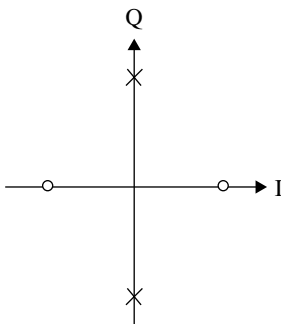
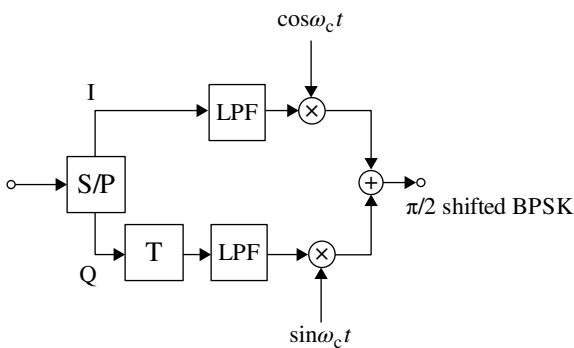


FIGURE 5.4 Block diagram of QPSK modulator.

FIGURE 5.5 Signal constellation of  $\pi/2$  shifted BPSK.FIGURE 5.6 Block diagram of  $\pi/2$  shifted BPSK modulator.

**5.3.1.4 Offset QPSK** If we decrease the bandwidth of the low-pass filter by one-half in Figure 5.6, we have an offset QPSK (OQPSK). OQPSK signal can be expressed as

$$s(t) = \sum_{n=-\infty}^{\infty} a_n h(t - nT_s) \cos \omega_c t + \sum_{n=-\infty}^{\infty} b_n h[t - (n + 1/2)T_s] \sin \omega_c t$$

If we let  $T_s = 2T$  in the above equation, this expression is equivalent to Equation 5.2. Thus, the difference between  $\pi/2$  shifted BPSK and OQPSK is only in the spectral bandwidth of the baseband signals. For example, assuming a non-return-to-zero (NRZ) signal, the impulse response becomes

$$h(t) = \begin{cases} 1 & 0 \leq t \leq T \\ 0 & \text{otherwise} \end{cases} \quad \text{for } \frac{\pi}{2} \text{ shifted BPSK}$$

and

$$h(t) = \begin{cases} 1 & 0 \leq t \leq 2T \\ 0 & \text{otherwise} \end{cases} \quad \text{for offset QPSK.}$$

The OQPSK signal phase never takes fixed points determined by the current symbol data. It scatters to different points owing to the interference between successive symbols. In this sense, the wording of “phase shift keying” may be inappropriate. Nevertheless, the demodulated signal with coherent detection shows no intersymbol interference since the in-phase and quadrature signals are orthogonal with each other. Figure 5.7 compares the waveforms of  $\pi/2$  shifted BPSK, QPSK, and OQPSK; NRZ signaling is assumed. The OQPSK signal can be obtained from  $\pi/2$  shifted BPSK signal, if the pulse duration is doubled. Similarly,  $\pi/2$  shifted BPSK signal can be obtained from OQPSK if RZ signaling with 50% duty ratio is applied instead of the NRZ signaling.

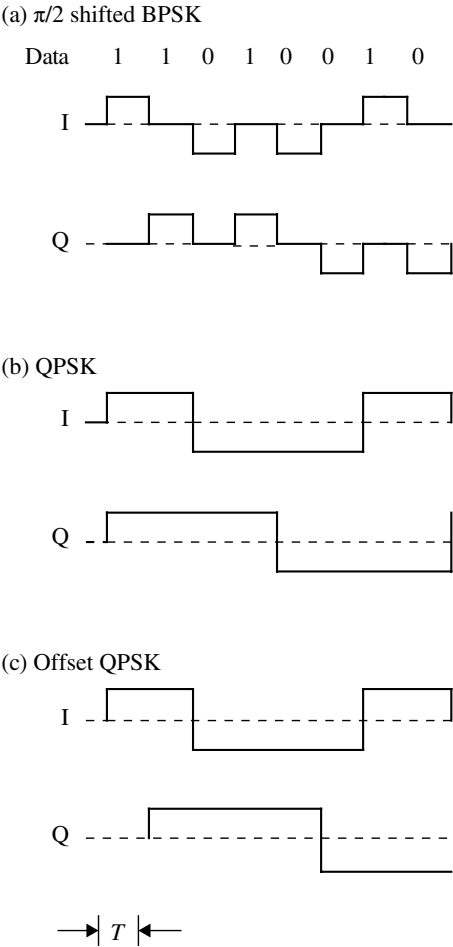
**5.3.1.5  $\pi/4$  Shifted QPSK** The  $\pi/4$  shifted QPSK signal can be generated by combining the output of two QPSK modulators, whose carrier phases are shifted by  $\pi/4$  radian (Fig. 5.8). The signal phase constellation is shown in Figure 5.9. The signal points are selected in turn from the two QPSK signal sets, which are marked with circles and crosses. The spectra of QPSK, OQPSK, and  $\pi/4$  shifted QPSK are the same with each other.

The  $\pi/4$  QPSK is described in Refs. [9–11]; in Ref. [10] it is called four-level differential phase exchange keying (DPEK) symmetric (phase exchange keying). The terminology “ $\pi/4$  shifted QPSK” is used in Ref. [11], where the signal is generated by phase shifting the QPSK signal by  $\pi/4$  for each symbol. For the  $\pi/4$  shifted QPSK, the fact that the signal phase change takes place at every symbol time guarantees to extract the clock-timing signal from it.

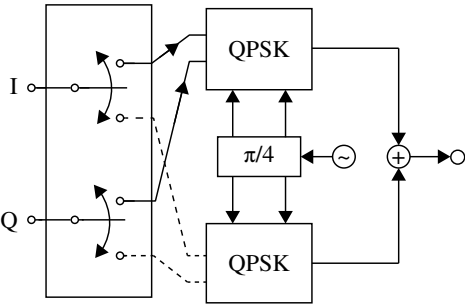
Differential encoding of transmit digital signal is widely used for PSK. The purpose of differential encoding is to avoid phase ambiguity in the carrier recovery or to introduce differential detection, as it is described later.

**5.3.1.6  $M$ -Ary PSK** This signal is expressed as

$$s(t) = \sum_{n=-\infty}^{\infty} A h(t - nT) \cos(\omega_c t + 2\pi a_n / M) \quad (a_n = 0, 1, \dots, M-1) \quad (5.3)$$



**FIGURE 5.7** Comparison of waveforms for (a)  $\pi/2$  shifted BPSK, (b) QPSK, and (c) offset QPSK.



**FIGURE 5.8** Block diagram of  $\pi/4$  shifted QPSK modulator.

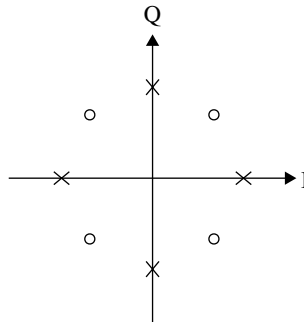


FIGURE 5.9 Signal constellation of  $\pi/4$  shifted QPSK.

### 5.3.2 Frequency Shift Keying

In this system, the instantaneous frequency is proportionally varied corresponding to the input digital signal. It is also called the digital FM. Some people differentiate FSK from digital FM in the following sense: they use the term “FSK” when NRZ signaling is adopted for the input digital signal and term “digital FM” if the input digital signal is band limited. An FSK signal, whose phase is continuous is called a continuous phase FSK (CPFSK). Switching the output from the two oscillators having different frequencies generates a discontinuous phase FSK signal. The phase discontinuity in this system causes the spectrum spreading. For this reason, discontinuous phase FSK is seldom used for mobile communication. CPFSK should be understood throughout this book unless stated otherwise.

An FSK signal has a constant envelope. Its spectrum is wider than the spectrum of a linear modulated signal. The shape of the spectrum is not the same as that of the baseband signal and it changes according to the modulation index as well as to the baseband signal. For example, Figure 5.10 shows spectra of digital FM with the modulation index as a parameter. When the modulation index is small enough, the spectrum becomes similar to that of the double sideband AM with the carrier signal component. When the modulation index is high, the spectrum has two peaks that correspond to the maximum frequency deviation. It is worthy to note that the spectra of an FSK signal can be controlled by band-limiting the input baseband signal: out-of-band radiation is effectively suppressed. This fact becomes important for digital mobile communications, where low out-of-band radiation is required.

The empirical formula for the bandwidth of an analog FM signal is well known as Carson’s bandwidth:

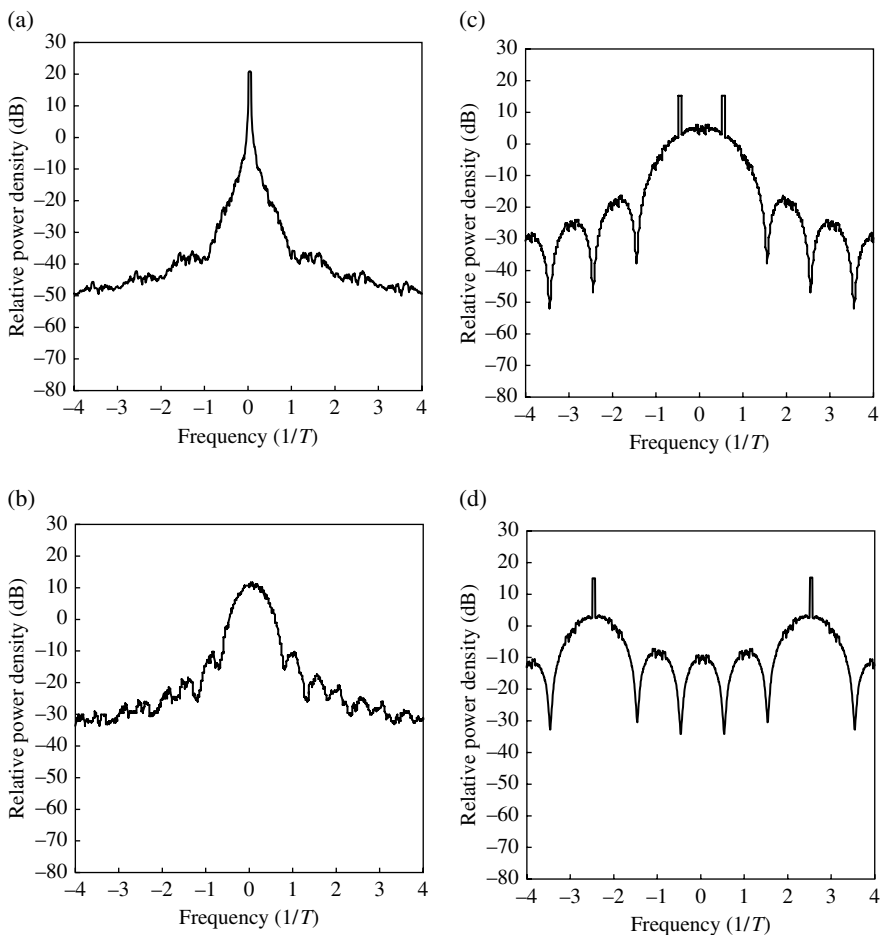
$$W = 2(\Delta F_m + f_b) = 2(m + 1)f_b$$

where  $\Delta F_m$  is the maximum frequency deviation,  $f_b$  the highest frequency of the baseband signal, and  $m(\equiv \Delta F_m/f_b)$  the modulation index. The Carson’s bandwidth shows that a bandwidth at least twice of the baseband bandwidth is required for an FM signal.

For the FSK signal, the modulation index is usually defined as

$$m = \frac{\Delta F_m}{1/(2T)} = 2\Delta F_m T \quad (5.4)$$





**FIGURE 5.10** FSK signal spectra for different modulation indices. (a)  $m=0.01$ , (b)  $m=0.5$ , (c)  $m=1$ , and (d)  $m=5$ .

where  $T$  is the symbol duration. Sometimes different definitions and notations are used for the modulation index such as  $h$ .

The phase of an FSK signal is given as

$$\begin{aligned}
 \varphi(t) &= \int_{-\infty}^t \omega(\tau) d\tau \\
 &= \int_{-\infty}^t \sum_{n=-\infty}^{\infty} k_F a_n h(\tau - nT) d\tau \\
 &= \sum_{n=-\infty}^{\infty} k_F a_n q(t - nT)
 \end{aligned} \tag{5.5}$$

where  $q(t) = \int_{-\infty}^t h(\tau) d\tau$ ,  $k_F$  is a proportional constant,  $a_n$  takes one of the discrete values corresponding to the input digital signal (e.g.,  $a_n = \pm 1$  for the 2-level FSK), and  $h(t)$  is the impulse response of the baseband filter. The actual maximum frequency deviation  $\Delta F_m = \max \{ \omega(t)/2\pi \}$  depends on the impulse response, as well as on data patterns. Instead of the actual  $\Delta F_m$ , we use

$$\Delta F_m \equiv \frac{k_F a_{\max}}{2\pi} \quad (5.6)$$

where  $a_{\max}$  is the highest of the discrete values.  $h(t)$  is normalized so that

$$\int_{-\infty}^{\infty} h(t) dt = T \quad (5.7)$$

The maximum phase shift caused by a symbol pulse becomes  $2\pi\Delta F_m T = m\pi$ . Equations 5.6 and 5.7 are valid for NRZ signaling, that is,  $h(t) = 1$  for  $|t| \leq T/2$  and  $h(t) = 0$  for  $|t| > T/2$ .  $\int_{-\infty}^{\infty} h(t) dt$  is the value of the transfer function  $H(\omega)$  at the dc-frequency, that is,  $H(0) = \int_{-\infty}^{\infty} h(t) dt$ . In case of  $H(0) = 0$ , for example in the bipolar or the bipolar coded (Section 3.2.6) FM system, we cannot use the definition of the modulation index: a specific definition must be made.

When the impulse response  $h(t)$  satisfies Nyquist's third criterion (Eq. 3.12), from Equations 5.4–5.7, we have

$$\begin{aligned} \varphi(iT) &= \sum_{n=-\infty}^i k_F a_n \int_{nT-T/2}^{nT+T/2} h(t-nT) dt \\ &= k_F T \sum_{n=-\infty}^i a_n \\ &= \pi m \sum_{n=-\infty}^i \frac{a_n}{a_{\max}} \end{aligned} \quad (5.8)$$

Nyquist's third criterion (Section 3.1.3) ensures that the phase shift of the FSK during a symbol time (i.e.,  $\Delta\varphi(iT) = \varphi(iT+T) - \varphi(iT)$ ) is determined by that symbol only: there is no intersymbol interference in  $\Delta\varphi(iT)$ .

With a special value of  $m$ , the phase  $\varphi(iT)$  takes some fixed points. Consider a 2-level FSK ( $a_n/a_{\max} = \pm 1$ ) with  $m = 0.5$ ; in this case, we get the signal phase constellation of  $\pi/2$  shifted BPSK. For a 4-level FSK ( $a_n/a_{\max} = \pm 1/3, \pm 1$ ) with  $m = 3/4$ , we get the signal phase constellation of  $\pi/4$  shifted QPSK.

### 5.3.3 Constant Envelope PSK

A constant envelope PSK signal can be obtained by removing the time integration in Equation 5.5:  $q(t)$  instead of  $h(t)$  represents the impulse response of the baseband filter.

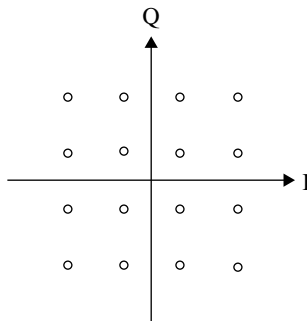


FIGURE 5.11 Signal constellation of 16QAM.

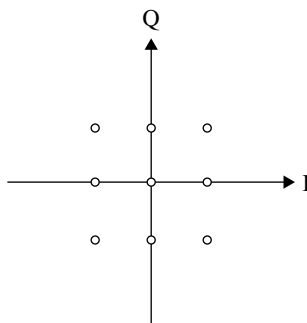


FIGURE 5.12 Signal constellation of duobinary 4QAM.

### 5.3.4 Quadrature Amplitude Modulation

In this system, the in-phase and quadrature components  $x(t)$  and  $y(t)$ , respectively, are amplitude-modulated independently from each other. QPSK can be also understood as QAM. A 4-level signal for each  $x(t)$  and  $y(t)$  results in  $4 \times 4$  signal levels. The signal constellation for 16-QAM is shown in Figure 5.11. When we apply the duobinary partial response (Section 3.2.6) to  $x(t)$  and  $y(t)$ , we get the  $3 \times 3$  signal constellation as shown in Figure 5.12. A higher level QAM such as 256-QAM was developed for digital transmission in public telephone channels or microwave channels.

## 5.4 POWER SPECTRAL DENSITY OF DIGITALLY MODULATED SIGNALS

The power spectral density (PSD) of a signal  $z(t)$  is defined as

$$s_z(\omega) = \left\langle \lim_{T_m \rightarrow \infty} \frac{1}{2T_m} \left| \int_{-T_m}^{T_m} z_T(t) e^{-j\omega t} dt \right|^2 \right\rangle$$

or

$$= \int_{-\infty}^{\infty} R_z(\tau) e^{-j\omega\tau} d\tau$$

where  $\langle \cdot \rangle$  means an ensemble average,  $z_T(t)$  is defined over the time range  $|t| < T_m$ , and  $R_z(\tau)$  is the autocorrelation function of  $z_T(t)$ ,

$$R_z(\tau) = \left\langle \lim_{T_m \rightarrow \infty} \frac{1}{2T_m} \int_{-T_m}^{T_m} z_T(t) z_T(t+\tau) dt \right\rangle$$

For a modulated signal, we have  $z_T(t) = \text{Re}[f_b(t)e^{j\omega_c t}]$ . Therefore, the autocorrelation function  $R_z(\tau)$  can be written as

$$R_z(\tau) = \frac{1}{2} \text{Re}[R_b(\tau)e^{j\omega_c \tau}] \quad (5.9)$$

where  $R_b(\tau)$  is the autocorrelation function of the complex baseband signal  $f_b(t)$  and is defined as

$$R_b(\tau) = \left\langle \lim_{T_m \rightarrow \infty} \frac{1}{2T_m} \int_{-T_m}^{T_m} f_b(t+\tau) f_b^*(t) dt \right\rangle$$

It can be shown that

$$R_b^*(\tau) = R_b(-\tau)$$

With the above relation, Equation 5.9 can be rewritten as

$$R_z(\tau) = \frac{1}{4} [R_b(\tau)e^{j\omega_c \tau} + R_b(-\tau)e^{-j\omega_c \tau}]$$

Then, the PSD of the modulated signal  $z_T(t)$  becomes

$$\begin{aligned} S_z(\omega) &= \frac{1}{4} \left[ \int_{-\infty}^{\infty} R_b(\tau) e^{-j(\omega - \omega_c)\tau} d\tau + \int_{-\infty}^{\infty} R_b(\tau) e^{j(\omega + \omega_c)\tau} d\tau \right] \\ &= \frac{1}{4} [S_b(\omega - \omega_c) + S_b(-\omega - \omega_c)] \end{aligned} \quad (5.10a)$$

If  $f_b(t)$  is real, we have (Eq. 2.37)

$$S_z(\omega) = \frac{1}{4} [S_b(\omega - \omega_c) + S_b(\omega + \omega_c)] \quad (5.10b)$$

where  $S_b(\omega)$  is the PSD of the complex baseband signal. Thus, we can see that the PSD of the complex baseband signal is shifted by  $\pm\omega_c$  due to the modulation.

### 5.4.1 Linear Modulation

For a linear modulation  $z_T(t)$  can be expressed generally as

$$\begin{aligned} z_T(t) &= x(t) \cos \omega_c t - y(t) \sin \omega_c t \\ &= \operatorname{Re} \left\{ [x(t) + jy(t)] e^{j\omega_c t} \right\} \quad (-T_m \leq t \leq T_m) \end{aligned}$$

where  $x(t)$  and  $y(t)$  are expressed as

$$\begin{aligned} x(t) &= \sum_n a_n h(t - nT), \\ y(t) &= \sum_n b_n h(t - nT) \end{aligned}$$

and  $a_n$  and  $b_n$  take discrete levels corresponding to the input data signal and  $h(t)$  is the impulse response of the baseband filter. If we assume random data,  $x(t)$  and  $y(t)$  become independent random variables. Owing to the independence of  $x(t)$  and  $y(t)$ ,  $R_b(\tau)$  is given as

$$R_b(\tau) = R_x(\tau) + R_y(\tau) \quad (5.11)$$

where  $R_x(\tau)$  and  $R_y(\tau)$  are the autocorrelation functions of  $x(t)$  and  $y(t)$ , respectively. In the present system,  $R_x(\tau) = R_y(\tau) = R_0(\tau)$ . Using this relation, we have from Equations 5.10b and 5.11

$$\begin{aligned} S_z(\omega) &= \int_{-\infty}^{\infty} R_0(\tau) \left\{ e^{-j(\omega - \omega_c)\tau} + e^{-j(\omega + \omega_c)\tau} \right\} d\tau \\ &= \frac{1}{2} [S_0(\omega - \omega_c) + S_0(\omega + \omega_c)] \end{aligned}$$

where  $S_0(\omega) = \int_{-\infty}^{\infty} R_0(\tau) e^{-j\omega\tau} d\tau$  is the PSD of  $x(t)$  or  $y(t)$ . Thus, the PSD for a linear modulation is given by shifting the PSD of the baseband modulating signal to the carrier frequency. The PSD of the baseband modulating digital signal is given by Equation 2.61.

### 5.4.2 Digital FM

For digital FM, unlike the linear modulation, derivation of the PSD is not as easy. There is ample literature on the PSD for digital FM. In the literature, different approaches are taken for the derivation. The simplest one is given in Ref. [12].

Rewriting Equation 5.5, we get

$$\phi(t) = 2\pi h \sum_{i=-\infty}^{\infty} \alpha_i q(t - iT)$$

where  $h$  is the modulation index and symbol  $\alpha_i$  takes one of the following values:

$$\alpha_i = \pm 1, \pm 3, \dots, \pm(M-1)$$

with a priori probabilities,

$$p_i = \text{Prob}(\alpha_j = i) \quad i = -M+1, -M+3, \dots, M-1 \quad \text{for any integer } j$$

and

$$q(t) = \int_{-\infty}^t g(\tau) d\tau$$

where  $g(t)$  is the impulse response of the baseband filter. The impulse response  $g(t)$  is assumed to be causal and its duration is truncated within  $LT$  symbol durations; therefore,

$$q(t) = \begin{cases} 0 & (t \leq 0) \\ q(LT) & (t \geq LT) \end{cases}$$

Here, the modulation index  $h$  is selected so that the maximum absolute phase change caused by a pulse over the duration of  $LT$ , or equivalently the maximum phase change over any symbol period  $T$ , is  $(M-1)h\pi$  with the normalization  $q(LT) = 1/2$ .

Omitting the derivation to be followed, we have the final result

$$S(\omega) = 2\text{Re} \left\{ \int_0^{LT} R(\tau) e^{-j\omega\tau} d\tau + \frac{e^{-j\omega LT}}{1 - c_\alpha e^{-j\omega T}} \int_0^T R(\tau + LT) e^{-j\omega\tau} d\tau \right\}$$

where

$$\begin{aligned} R(\tau) &= R(\tau' + mT) \\ &= \frac{1}{T} \int_0^T \prod_{n=1-L}^{m+1} \left\{ \sum_{\substack{k=-(M-1) \\ k:\text{odd}}}^{M-1} p_k \exp \{ j2\pi h k [q(t + \tau' - (n-m)T) - q(t - nT)] \} \right\} dt \end{aligned}$$

and  $\tau$  is over the period  $0 - (L+1)T$ . In this case, the time difference  $\tau$  has been written as

$$\tau = \tau' + mT \quad (0 \leq \tau' \leq T, m = 0, 1, \dots)$$

and

$$c_\alpha = \sum_{\substack{k=-(M-1) \\ k:\text{odd}}}^{M-1} p_k \exp(j\pi h k)$$

A discrete component of the PSD appears when  $|c_\alpha| = 1$ .

## 5.5 DEMODULATION

Demodulation or detection is a process where the transmit baseband signal is recovered from the modulated signal. The methods of demodulation are classified into coherent or synchronous detection and noncoherent detection, depending on whether a carrier signal is or is not used, respectively. Noncoherent detection includes the envelope detection, differential detection, and frequency detection. Differential detection is also called differentially coherent detection or delay detection.

### 5.5.1 Coherent Detection

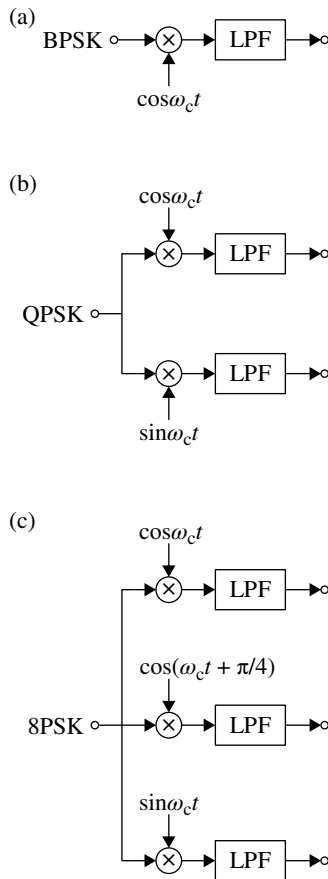
In coherent detection systems, the baseband modulating signal is obtained by multiplying by a carrier signal, which is synchronized to the modulated signal. This process is the inverse of linear modulation. Hence, coherent detection could be called linear demodulation.

Coherent demodulators for BPSK, QPSK, and 8PSK are shown in Figure 5.13a, b, and c, respectively; the decision circuit follows the demodulator. Coherent detection has a superior error rate compared to noncoherent detection under static conditions. A problem with the coherent detection for mobile radio channels is that the carrier recovery circuit becomes unstable due to low signal-to-noise power ratio and the random FM effect of fading. Due to the random FM effect, the bit error rate versus carrier to noise power ratio curve shows a floor. The floor error rate is called the irreducible error rate. The floor error rate is higher for coherent detection than for noncoherent detection.

**5.5.1.1 Carrier Recovery** In contrast to the analog transmission system with coherent demodulation, typically, in digital transmission systems, the carrier signal is not sent with the modulated signal, but it is regenerated from the modulated signal. Regeneration of the carrier signal component for digital modulation is possible because the modulated signal phase takes specific discrete phase points at each symbol time.

In the regeneration process, first a discrete spectrum component is obtained that is synchronized to the carrier signal through appropriate operations on the received signal. Later, the discrete spectrum component signal is applied to a tank circuit or a phase-locked loop (PLL) circuits which is tuned to the discrete spectrum component in order to reduce jitter of the carrier signal. The  $Q$ -value of the tank circuit or, equivalently, the loop bandwidth of the PLL circuit must be compromised with the SNR and rise-up speed of the carrier signal regeneration. In order to boost the rise-up speed, a preamble signal, which is rich in discrete spectrum components, is added before the information-carrying modulated signal. The exception to this is described in Ref. [13], where the received signal is stored and then demodulated.

The principles of the carrier signal regeneration methods fall into three categories: frequency multiplication, inverse modulation, and remodulation. In the frequency-multiplication method, modulated signal is subjected to a frequency-multiplying circuit followed by a tank circuit or a PLL circuit, and a frequency-dividing circuit.

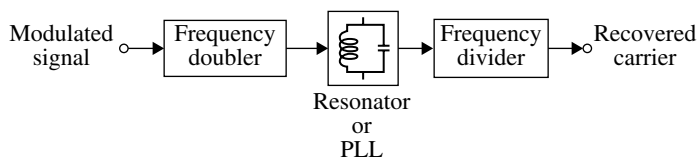


**FIGURE 5.13** Coherent demodulator: (a) BPSK, (b) QPSK, and (c) 8PSK.

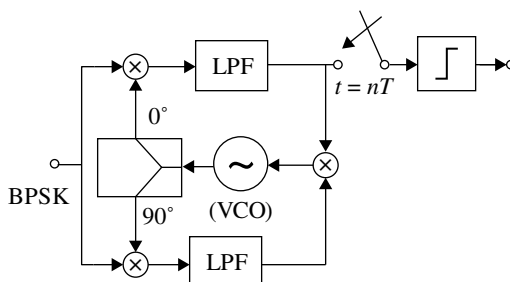
The block diagram of this method is shown in Figure 5.14. From the signal phase constellation (Fig. 5.2), we see that the double-, quadruple-, and octuple-frequency multiplication of the BPSK, QPSK, and 8PSK signals, respectively, result in a signal that takes one phase ( $m\pi$ ,  $m$ : integer) at each symbol time. Thus, the discrete spectrum component signal is obtained.

A method of coherent detection, where the frequency multiplication and division are equivalently performed at the baseband, is known as the Costas-loop demodulator (Fig. 5.15). Let the modulated signal be  $s(t) = 2A\cos[\omega_c t + \varphi(t)]$ , the outputs of the quadrature detector become  $x(t) = A\cos[\varphi(t) + \theta_0]$  and  $y(t) = A\sin[\varphi(t) + \theta_0]$ , where  $\theta_0$  is the phase error between the regenerated carrier signal and the modulated signal. After multiplying  $x(t)$  and  $y(t)$ , the input signal of the voltage controlled oscillator (VCO) becomes  $(A^2/2)\sin[2\varphi(t) + 2\theta_0]$ . This result shows a frequency-doubling at baseband. For the BPSK signal, the phase of the frequency-doubled signal  $2\varphi(t)$  takes a value of zero or  $2\pi$ . Therefore, the Costas-loop circuit has a negative feedback

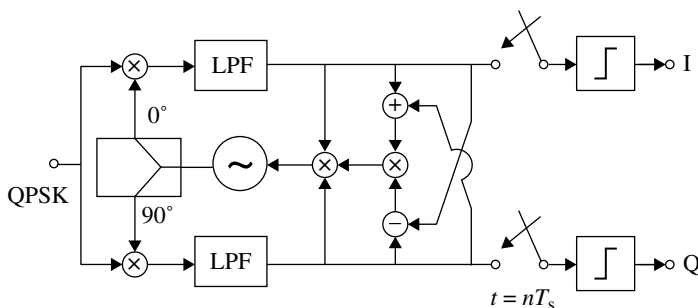




**FIGURE 5.14** Block diagram of carrier regeneration circuit with frequency-multiplying method.



**FIGURE 5.15** Costas-loop demodulator for BPSK signal.



**FIGURE 5.16** Costas-loop demodulator for QPSK signal.

control to achieve  $\theta_0 \rightarrow 0$ , and the synchronization of the recovered carrier signal to the modulated signal is established. The Costas-loop detection can be applied to the QPSK, by frequency-quadrupling at baseband (Fig. 5.16).

Block diagram of coherent detection with inverse modulation is shown in Figure 5.17. The input signal is inversely modulated with the baseband signal to remove the effect of modulation. The inverse modulation is equivalent to remodulation in BPSK system, since the signal phase takes values 0 or  $\pi$ . The recovered carrier signal is synchronized to the input signal through the use of a VCO.

The coherent detection with remodulation is shown in Figure 5.18. The replica of the input signal is produced by modulation of the recovered carrier signal (VCO output) with the demodulated baseband signal. Thus, the input modulated signal is tracked by the replica signal through the feedback control circuit including a VCO, a detector, and a modulator.

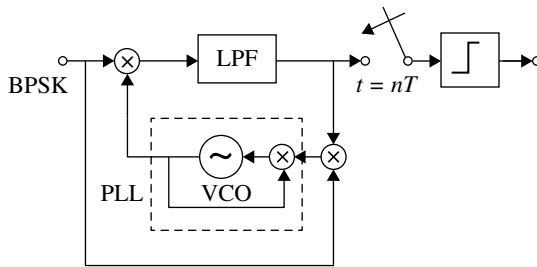


FIGURE 5.17 Coherent detection with inverse modulation.

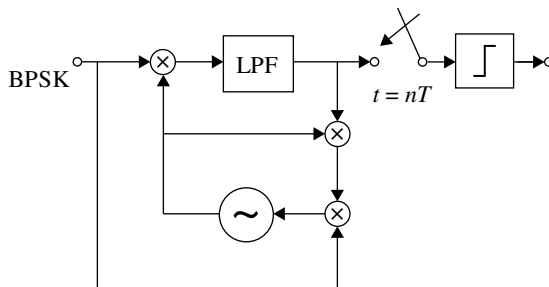


FIGURE 5.18 Coherent detection with remodulation.

Phase ambiguity can still remain in the recovered carrier. For example, an ambiguity of  $\pm\pi$  exists in a frequency-dividing ( $1/2$ ) circuit of the frequency-multiplication method, as well as in the Costas-loop circuit. In mobile radio communication, the carrier recovery circuit loses synchronization because of degradation in signal-to-noise power ratio or random FM effects of fast fading. Although the synchronization is automatically reestablished, phase ambiguity of the carrier may exist for each recovery of synchronization. For DPSK, where the information is conveyed in the relative phase change between two successive symbols phase, ambiguity can be removed by the differential encoding (Section 3.2.7). With DPSK, error propagation exists: one error in the detection causes two successive symbol errors. The error propagation is avoided through use of coherent detection with an absolute phase carrier. For this purpose, a pilot signal to determine the absolute phase must be transmitted either continuously or intermittently with the information signal.

**5.5.1.2 Noise Power at the Output of the Coherent Detector** We assume narrow-band bandpass noise,  $n(t) = n_x(t) \cos \omega_c t - n_y(t) \sin \omega_c t$ , with PSD of  $N_0/2$

$$S_n(\omega) = \begin{cases} N_0/2 & (|\omega - \omega_c| \leq \omega_a) \\ 0 & (\text{otherwise}) \end{cases} \quad (5.12)$$

where  $\omega_a$  is an arbitrary constant, which is high enough to cover the modulated signal bandwidth and is lower than the carrier frequency  $\omega_c$ .

Multiplying the pass-band noise by  $2\cos\omega_c t$  and  $2\sin\omega_c t$ , we have the in-phase and quadrature baseband noise components  $n_{xd}(t) = n_x(t)$  and  $n_{yd}(t) = n_y(t)$ . From Equations 2.84 and 5.12, PSD for the demodulated noise is

$$S_{xd}(\omega) = S_{yd}(\omega) \equiv S_d(\omega) = \begin{cases} N_0 & (-\omega_a < \omega < \omega_a) \\ 0 & (\text{otherwise}) \end{cases}$$

The noise power at the output of the receive baseband filter becomes

$$\begin{aligned} N &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_d(\omega) |G(\omega)|^2 d\omega \\ &= \frac{N_0}{2\pi} \int_{-\infty}^{\infty} |G(\omega)|^2 d\omega \end{aligned} \quad (5.13a)$$

or from Parseval's theorem (Eq. 2.43),

$$= N_0 \int_{-\infty}^{\infty} g^2(t) dt \quad (5.13b)$$

where  $G(\omega)$  is the transfer function of the filter and  $g(t) \leftrightarrow G(\omega)$ .

**5.5.1.3 Error Rate Analysis** The error rate analysis of coherent detection is easy, since it is a linear process. However, this is true only when the recovered carrier is ideal. If we take the carrier recovery process into consideration, the analysis becomes difficult, since the system includes nonlinear operations. Burst errors occur when the carrier recovery circuit loses synchronization. The ideal carrier is assumed in the following analyses.

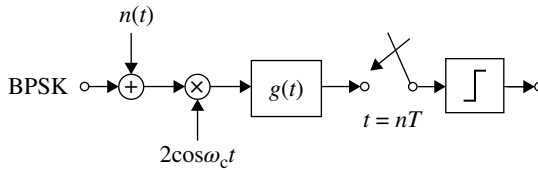
We assume symbol by symbol detection with a matched filter that meets Nyquist's first criterion (no intersymbol interference, Section 3.1.1) and white Gaussian noise. The baseband signal obtained by demodulation is decided at each symbol time.

**BPSK** Let us consider first the coherent detection of a BPSK signal (Fig. 5.19). The received signal is the sum of the BPSK signal and noise

$$\begin{aligned} r(t) &= s(t) + n(t) \\ &= \sum_{n=-\infty}^{\infty} a_n h(t - nT) \cos\omega_c t + n_x(t) \cos\omega_c t - n_y(t) \sin\omega_c t \quad (a_n = \pm A) \end{aligned}$$

Multiplying  $r(t)$  by the carrier signal  $2\cos\omega_c t$ , we have the demodulated signal

$$\begin{aligned} r_d(t) &= s'_d(t) + n_d(t) \\ &= \sum_{n=-\infty}^{\infty} a_n h(t - nT) + n_x(t) \end{aligned}$$



**FIGURE 5.19** Coherent detection of a BPSK signal.

The demodulated signal is applied to a receive filter with an impulse response  $g(t)$ . Let us denote the output signal level at the sampling instant  $t_0$  by  $\pm d/2$ , that is,

$$s_d(t_0) = \pm Ah(t) * g(t) \big|_{t=t_0} = \pm d/2$$

A decision error occurs for the symbol  $m$  when  $s_d(t_0)$  deviates by more than  $d/2$  owing to the noise. The noise power  $N = \sigma_n^2$  is given by Equation 5.13a or 5.13b. Hence, from Equation 3.27 with  $d_m = d/2$ ,  $P(\pm M) = 1$  and considering Equation 2.68, the error rate is given as

$$P_e = \frac{1}{2} \operatorname{erfc} \left( \frac{d}{2\sqrt{2}\sigma_n} \right) = Q \left( \frac{d}{2\sigma_n} \right) \quad (5.14a)$$

Rewriting this equation using the signal-to-noise power ratio at the sampling instant  $S/N = d^2/4\sigma_n^2$ , we have

$$P_e = \frac{1}{2} \operatorname{erfc} \left( \sqrt{\frac{1}{2} \frac{S}{N}} \right) = Q \left( \sqrt{\frac{S}{N}} \right) \quad (5.14b)$$

When the receive filter is the matched filter, we have  $g(t) = h(t_0 - t)$  (Eq. 3.30). Then from Equation 3.33, we have

$$s_d(t_0) = \pm \frac{d}{2} = \pm A \int_{-\infty}^{\infty} h^2(t) dt$$

and from Equation 5.13b

$$N = \sigma_n^2 = N_0 \int_{-\infty}^{\infty} h^2(t) dt \quad (5.15)$$

Hence we have

$$\begin{aligned} \frac{S}{N} &= \frac{s_d^2(t_0)}{\sigma_n^2} \\ &= \frac{\left[ A \int_{-\infty}^{\infty} h^2(t) dt \right]^2}{N_0 \int_{-\infty}^{\infty} h^2(t) dt} = \frac{A^2}{N_0} \int_{-\infty}^{\infty} h^2(t) dt \end{aligned} \quad (5.16)$$

The average energy per symbol (or per bit) for the BPSK is given as

$$E_s = E_b = \frac{A^2}{2} \int_{-\infty}^{\infty} h^2(t) dt \quad (5.17a)$$

or using Parseval's theorem

$$= \frac{A^2}{2} \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega \quad (5.17b)$$

Inserting Equation 5.17a into Equation 5.16, we have  $\frac{S}{N} = 2 \frac{E_b (= E_s)}{N_0}$ .  
Hence we obtain

$$P_e = \frac{1}{2} \text{erfc}(\sqrt{\lambda}) \quad (5.18a)$$

$$= Q(\sqrt{2\lambda}) \quad (5.18b)$$

where  $\lambda = E_s/N_0 = E_b/N_0$ .

Error rates for BPSK and other techniques are shown in Figure 5.20.

**QPSK** The QPSK signal is given by Equation 5.1. The average energy per symbol can be calculated as

$$E_s = A^2 \int_{-\infty}^{\infty} h^2(t) dt \quad (5.19)$$

With a discussion similar to the analysis of the BPSK, we obtain the signal-to-noise power ratio at the output of the matched filter at the sampling instant for both the in-phase and quadrature signals as

$$\frac{S}{N} = \frac{A^2}{N_0} \int_{-\infty}^{\infty} h^2(t) dt$$

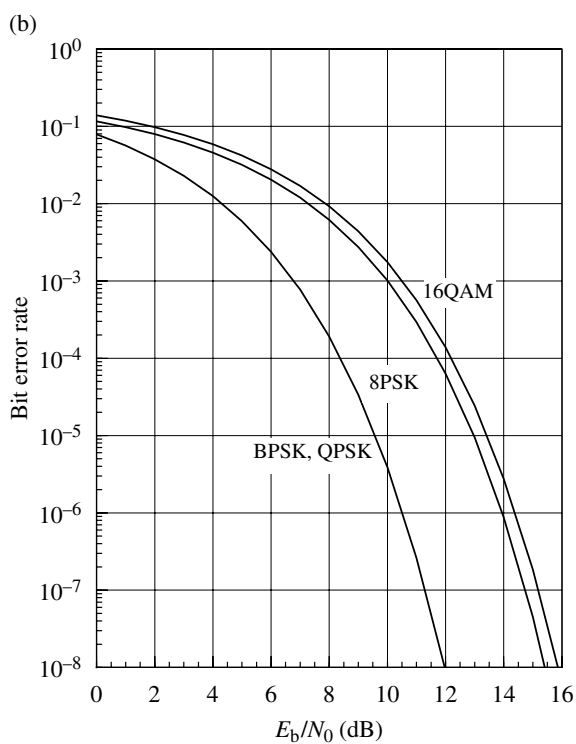
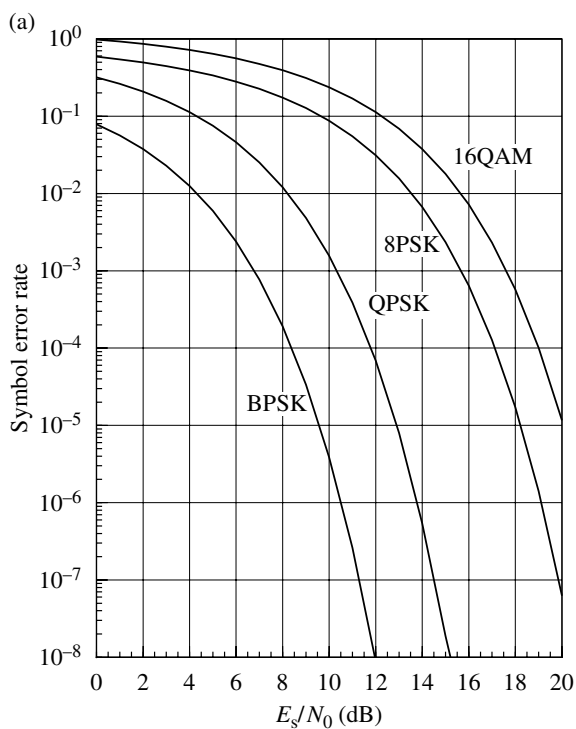
which is the same as that of the BPSK system. Rewriting the above equation with Equation 5.19, we have

$$\frac{S}{N} = \frac{E_s}{N_0} (\equiv \lambda)$$

Using the relation  $E_s = 2E_b$ , we get

$$\frac{S}{N} = 2 \frac{E_b}{N_0}$$

A symbol error occurs when there is a decision error in the in-phase signal or quadrature signal. The error probability  $q$  for the in-phase or quadrature signal is



**FIGURE 5.20** Error rates. (a) Symbol error rate vs. symbol energy to noise power density ratio and (b) bit error rate versus bit energy to noise power density ratio.

$q = \frac{1}{2} \operatorname{erfc}(\sqrt{\lambda}/2)$ . Since the in-phase and quadrature noise are uncorrelated, the probability that a symbol is correctly decided becomes  $(1-q)^2$ . Thus, we have the symbol error rate as

$$\begin{aligned} P_{\text{es}} &= 1 - (1-q)^2 \\ &= 2q - q^2 \end{aligned} \quad (5.20a)$$

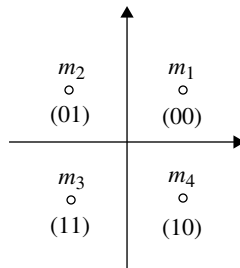
$$\approx 2q \quad (5.20b)$$

where  $q = \frac{1}{2} \operatorname{erfc}(\sqrt{\lambda}/2) = Q(\sqrt{\lambda}) = \frac{1}{2} \operatorname{erfc}(\sqrt{E_b/N_0})$  and  $\lambda = E_s/N_0$  is the energy per symbol to noise power density ratio. The bit error rate depends on the assignment of the bit pairs of the four phase points. When we use a Gray code assignment (Fig. 5.21), a symbol error  $m_1 \rightarrow m_2$  or  $m_1 \rightarrow m_4$  causes one error in the two bits and a symbol error  $m_1 \rightarrow m_3$  causes a two-bit error. Hence, we have

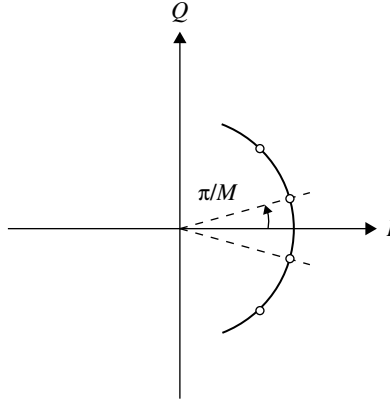
$$\begin{aligned} P_{\text{eb}} &= \frac{1}{2}q(1-q) + \frac{1}{2}q(1-q) + q^2 = q \\ &= \frac{1}{2} \operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) \end{aligned} \quad (5.21)$$

Thus, bit error rate as a function of  $E_b/N_0$  is the same for BPSK and QPSK with the Gray code.

OQPSK and  $\pi/4$  shifted QPSK have the same error rate as QPSK. This can be seen from the fact that the in-phase and quadrature signals are independent from each other, and decisions are made symbol by symbol using a matched filter. Since no intersymbol interference is considered, a demodulated pulse signal in the in-phase and quadrature components can be decided independently. Thus, the timing offset between the in-phase and quadrature components in the OQPSK and the  $\pi/4$  phase shift in the  $\pi/4$  shifted QPSK have no effect on the decision-making process. Therefore, the signal-to-noise power ratio at the sampling instant at the output of the matched filter shows the same value for QPSK, OQPSK, and  $\pi/4$  shifted QPSK.



**FIGURE 5.21** The Gray-coded QPSK.



**FIGURE 5.22**  $M$ -ary PSK signals.

Symbol and bit error rates for QPSK given with Equations 5.20b and 5.21 are shown in Figure 5.19a and b, respectively.

**$M$ -Ary PSK** We consider a PSK signal whose phase takes  $M$  equidistant points (Fig. 5.22) [1]. A decision error occurs when the received signal phase deviates more than  $\pi/M$  due to the noise. The probability density function of the phase of a sinusoidal signal plus narrow-band bandpass Gaussian noise is given by Equation 2.86. Then the (symbol) error rate can be calculated as

$$\begin{aligned}
 P_{\text{es}} &= 1 - \int_{-\pi/M}^{\pi/M} p(\theta) d\theta \\
 &= 1 - \frac{1}{2\pi} \int_{-\pi/M}^{\pi/M} e^{-\lambda} \left\{ 1 + \sqrt{4\pi\lambda} \cos\theta e^{\lambda \cos^2\theta} \left[ 1 - Q\left(\sqrt{2\lambda} \cos\theta\right) \right] \right\} d\theta
 \end{aligned} \tag{5.22}$$

where  $\lambda$  is the pulse energy (energy per symbol) to the noise power density ratio at the passband. For  $\lambda \gg 1$  and  $M \gg 2$ , using  $Q(x) \approx \left( \frac{1}{x\sqrt{2\pi}} \right) e^{-\frac{x^2}{2}}$  ( $x \gg 1$ ), Equation 5.22 is approximated as

$$P_{\text{es}} \approx 2Q\left(\sqrt{2\lambda \sin^2 \frac{\pi}{M}}\right) \tag{5.23a}$$

$$\approx 2Q\left(\sqrt{\frac{2\pi^2 \lambda}{M^2}}\right) \tag{5.23b}$$

For  $\lambda \gg 1$ , the probability that the phase deviates by more than  $2\pi/M$  can be neglected. If we use Gray coding, we expect that the bit error rate  $P_{\text{eb}}$  becomes  $P_{\text{es}}/\log_2 M$ . For  $M=2$  and  $M=4$ , Equation 5.22 should become Equation 5.18a or b and Equation 5.21, respectively.



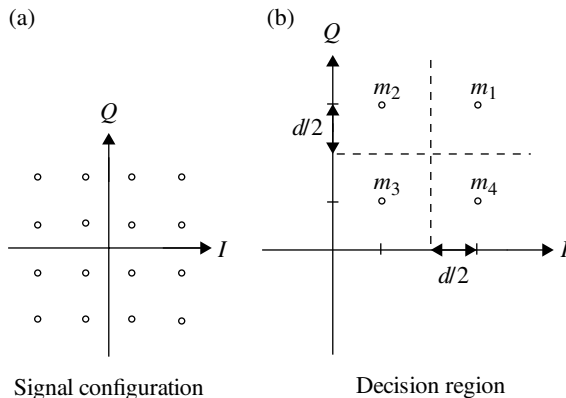


FIGURE 5.23 16QAM.

Symbol and bit error rates given by Equation 5.23a are shown for 8PSK in Figure 5.20a and b, respectively.

**16QAM** Let us consider 16QAM [1]. The signal configuration and decision regions are shown in Figure 5.23a and b, respectively. From Figure 5.23b, the probability that the symbol  $m_1$  is correctly decided becomes

$$\begin{aligned}
 P(C | m_1) &= \text{prob}\left(n_x > -\frac{d}{2}\right) \text{prob}\left(n_y > -\frac{d}{2}\right) \\
 &= \left[1 - Q\left(\frac{d/2}{\sigma_n}\right)\right]^2
 \end{aligned}$$

where  $\sigma_n^2$  is the noise power at the output of the baseband filter.

Assuming  $p = 1 - Q((d/2)/\sigma_n)$  and using previous arguments, we get

$$\begin{aligned}
 P(C | m_1) &= p^2 \\
 P(C | m_2) &= P(C | m_4) = p(2p - 1) \\
 P(C | m_3) &= (2p - 1)^2
 \end{aligned}$$

Assuming equal occurrence of each symbol and using the symmetry of the signal configuration, the probability of correct decision becomes

$$P(C) = \frac{1}{16} \sum_{i=1}^{16} P(C | m_i) = \left(\frac{3p-1}{2}\right)^2$$

Hence, the symbol error probability is given as

$$P_{\text{es}} = 1 - P(C) = \frac{9}{4} \left(p + \frac{1}{3}\right) (1 - p)$$

For  $p \approx 1$ , we have

$$P_{\text{es}} \approx 3(1-p) = 3Q \left( \frac{d}{2\sigma_n} \right)$$

At the sampling instant, the signal power of the symbol  $m_1$  is

$$S_1 = \left( \frac{3d}{2} \right)^2 + \left( \frac{3d}{2} \right)^2 = \frac{9}{2} d^2$$

Similarly, we have

$$S_2 = S_4 = \frac{5}{2} d^2, \quad S_3 = \frac{1}{2} d^2$$

Hence, the average power  $\bar{S}$  becomes

$$\bar{S} = \frac{1}{4} (S_1 + S_2 + S_3 + S_4) = \frac{5}{2} d^2 \quad (5.24)$$

Thus,

$$P_{\text{es}} \approx 3Q \left( \sqrt{\frac{\bar{S}}{5N}} \right) \quad (5.25)$$

where  $N = 2\sigma_n^2$ .

In the following, the average  $\bar{S}/N$  is related to the input signal energy per symbol to the noise PSD of  $N_0/2$ . The 16QAM signal can be expressed as

$$s(t) = \sum_{n=-\infty}^{\infty} a_n h(t-nT) \cos \omega_c t + \sum_{n=-\infty}^{\infty} b_n h(t-nT) \sin \omega_c t \quad (5.26)$$

where  $a_n$  and  $b_n$  take a value of  $\pm A, \pm 3A$ . Multiplying Equation 5.26 by  $2\cos \omega_c t$  and  $2\sin \omega_c t$ , we have the demodulated in-phase and quadrature signals:

$$s_{\text{dx}}(t) = \sum_{n=-\infty}^{\infty} a_n h(t-nT),$$

$$s_{\text{dy}}(t) = \sum_{n=-\infty}^{\infty} b_n h(t-nT)$$

Applying these signals to the matched filter, we have a signal level at the sampling instant as

$$A \int_{-\infty}^{\infty} h^2(t) dt = \frac{d}{2}$$

The average symbol energy  $E_s$  is given as

$$\begin{aligned} E_s &= \frac{1}{4} \left[ A^2 + (-A)^2 + (3A)^2 + (-3A)^2 \right] \times \int_{-\infty}^{\infty} h^2(t) dt \\ &= 5A^2 \int_{-\infty}^{\infty} h^2(t) dt \end{aligned} \quad (5.27)$$

From Equations 5.24 and 5.15, we have

$$\frac{\bar{S}}{N} = \frac{\frac{5}{2} \left[ 2A \int_{-\infty}^{\infty} h^2(t) dt \right]^2}{2N_0 \int_{-\infty}^{\infty} h^2(t) dt} = \frac{5A^2}{N_0} \int_{-\infty}^{\infty} h^2(t) dt$$

Inserting Equation 5.27 into the above equation, we get

$$\frac{\bar{S}}{N} = \frac{E_s}{N_0} = \frac{4E_b}{N_0} \quad (5.28)$$

where we used  $E = E_s/4$  (Eq. 2.59).

Hence from Equations 5.25 and 5.28, we have

$$P_{es} \approx 3Q \left( \sqrt{\frac{E_s}{5N_0}} \right) = 3Q \left( \sqrt{\frac{4E_b}{5N_0}} \right) \quad (5.29)$$

Assuming Gray coding, the bit error rates  $P_{eb}$  becomes  $P_{eb} \approx P_{es}/4$ . Symbol and bit error rate performances for 16QAM given by Equation 5.29 are shown in Figure 5.19a and b, respectively.

### 5.5.2 Envelope Detection

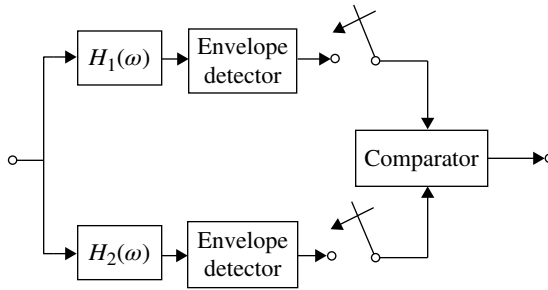
In a narrow sense, noncoherent detection means envelope detection. The detector consists of the matched filter followed by the envelope detector, a sampler and a decision devices (Fig. 5.24). This detection is optimal when the phase difference between the input signal carrier and the recovered carrier is unknown. Let us assume that envelopes of modulated signals are orthogonal to each other for different symbol data. Thus, both signal and noise appear at the output of the matched filter while only noise appears at the output of the other filter.

Using the probability density functions of envelopes of the sinusoidal signal plus noise (Eq. 2.85) and the noise (Eq. 2.79), the error rate can be analyzed.

For an ASK signal, where a modulated signal is expressed as  $s(t) = \sum_{n=-\infty}^{\infty} a_n h(t - nT) \cos(\omega_c t + \theta_0)$  ( $a_n = A$  or  $0$ ), the error rate is given in [1] as

$$P_e \approx \frac{1}{2} e^{-\lambda/2} \quad (\lambda \gg 1) \quad (5.30)$$

where  $\lambda \equiv E_b/N_0$  is the energy per bit to noise power density.



**FIGURE 5.24** Noncoherent detection.

For an FSK signal, the error rate is given in [1] as

$$P_e = \frac{1}{2} e^{-\lambda/2} \quad (5.31)$$

### 5.5.3 Differential Detection

The phase difference between two symbols can be detected. Detection of the phase difference is carried out by multiplying the preceding signal, which is delayed by a symbol time, with the present signal. Figure 5.25a and b shows the differential detection systems for BPSK and QPSK, respectively. It is called alternatively, as differentially coherent detection, delay detection or phase comparison detection. Here  $M$ -ary PSK is considered.

The operation of the differential detection (Fig. 5.25) is analyzed in the following. The  $M$ -ary PSK signal is given by Equation 5.3. Denote the output signal of the matched filter as

$$s_m(t) = \sum_{n=-\infty}^{\infty} A g(t-nT) \cos\left(\omega_c t + \frac{2\pi a_n}{M}\right) (a_n = 0, 1, 2, \dots, M-1)$$

where  $g(t) = h(t) * h(t_0 - t)$ .

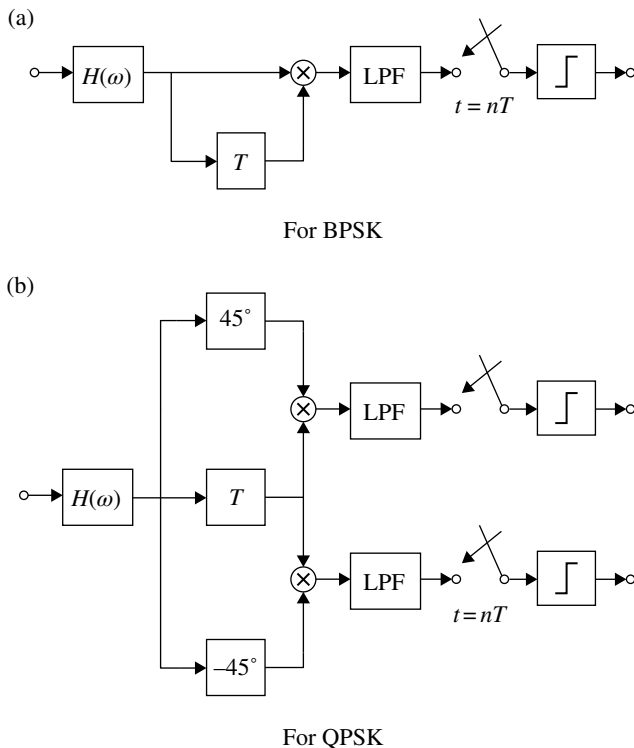
The output signal of the delay line becomes

$$s_d(t) = \sum_{n=-\infty}^{\infty} A g(t-nT) \cos\left[\omega_c (t-T) + \frac{2\pi a_{n-1}}{M}\right]$$

Then for BPSK ( $M=2$ ), the output signal of the low-pass filter is

$$s_o(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{1}{2} A^2 g(t-mT) g(t-nT) \cos[\omega_c T + (a_m - a_{n-1})\pi] (a_n = 0 \text{ or } 1)$$

where it is assumed that the low-pass filter attenuates the higher harmonics and its bandwidth is wide enough to pass the baseband signal.



**FIGURE 5.25** Differential detection system.

If  $g(nT)=0$  for  $n \neq 0$ , that is, the system satisfies Nyquist's first criterion, we have at the sampling instant

$$s_o(nT) = \frac{1}{2} A^2 g^2(0) \cos[\omega_c T + (a_n - a_{n-1})\pi]$$

When  $\omega_c T = 2n\pi$  ( $n = 1, 2, \dots$ ), we have

$$s_o(nT) = \frac{1}{2} A^2 g^2(0) \cos[(a_n - a_{n-1})\pi]$$

then

$$s_o(nT) \begin{cases} > 0 & (a_n = a_{n-1}) \\ < 0 & (a_n \neq a_{n-1}) \end{cases}$$

We can detect whether a data change has or has not occurred between the symbols by determining the polarity of  $s_o(nT)$ . Since we detect the data change, the data must be differentially encoded (Section 3.2.7) at the transmitter.

For QPSK, the demodulated signals can be expressed as

$$\begin{aligned} s_{0x}(nT) &= A \cos \left[ \frac{(a_n - a_{n-1})\pi}{2} + \frac{\pi}{4} \right] \\ s_{0y}(nT) &= A \sin \left[ \frac{(a_n - a_{n-1})\pi}{2} + \frac{\pi}{4} \right] \quad (a_n = 0, 1, 2, 3) \end{aligned}$$

The phase difference  $(a_n - a_{n-1})\pi/2 \pmod{2\pi}$  becomes 0,  $\pm\pi/2$ , and  $\pi$  and can be determined by a threshold detection on  $s_{0x}(nT)$  and  $s_{0y}(nT)$ .

**Error Rate** Consider a signal and noise at  $t=t_1$  and  $t_2$  as

$$\begin{aligned} s(t_1) &= A \cos[\omega_c t_1 + \varphi(t_1)] + n_x(t_1) \cos[\omega_c t_1 + \varphi(t_1)] \\ &\quad - n_y(t_1) \sin[\omega_c t_1 + \varphi(t_1)] \\ s(t_2) &= A \cos[\omega_c t_2 + \varphi(t_2)] + n_x(t_2) \cos[\omega_c t_2 + \varphi(t_2)] \\ &\quad - n_y(t_2) \sin[\omega_c t_2 + \varphi(t_2)] \end{aligned}$$

The phase difference between  $s(t_1)$  and  $s(t_2)$  is

$$\begin{aligned} \angle s(t_2) - \angle s(t_1) &= \omega_c(t_2 - t_1) + \varphi(t_2) - \varphi(t_1) \\ &\quad + \tan^{-1} \frac{n_y(t_1)}{A + n_x(t_1)} - \tan^{-1} \frac{n_y(t_2)}{A + n_x(t_2)} \end{aligned}$$

where the term  $\varphi(t_2) - \varphi(t_1)$  expresses the phase shift with modulation and the term  $\psi \equiv \tan^{-1} \frac{n_y(t_1)}{A + n_x(t_1)} - \tan^{-1} \frac{n_y(t_2)}{A + n_x(t_2)}$  corresponds to the deviation due to noise. The probability density function of the phase  $\theta$  for the signal and noise is given by Equation 2.86. If the noise at  $t=t_1$  and  $t_2$  are uncorrelated, the probability density of the phase difference  $\psi$  is given as

$$p(\psi) = \int_{-\pi}^{\pi} p(\theta_1) p(\theta_1 + \psi) d\theta_1 \quad (-\pi \leq \psi \leq \pi)$$

It is shown by Pawula, Rice, and Roberts [14] as follows:

$$\text{Prob}\{\psi_1 \leq \psi \leq \psi_2\} = F(\psi_2) - F(\psi_1) \quad (5.32)$$

where

$$F(\psi) = -\frac{\sin \psi}{4\pi} \int_{-\pi/2}^{\pi/2} \frac{e^{-\lambda(1 - \cos \psi \cos t)}}{1 - \cos \psi \cos t} dt \quad (5.33)$$

and  $\lambda = A^2/2\sigma_n^2$  is the average signal-to-noise power ratio at the sampling instants. When a matched filter receiver is considered,  $\lambda$  becomes energy per symbol-to-noise density ratio.

A symbol error occurs when  $|\psi| > \pi/M$ . Thus, the symbol error rate is given as

$$\begin{aligned} P_e &= \int_{\pi/M}^{\pi} p(\psi) d\psi + \int_{-\pi}^{-\pi/M} p(\psi) d\psi \\ &= 2 \int_{\pi/M}^{\pi} p(\psi) d\psi = 2 \text{Prob}\{\pi/M \leq \psi \leq \pi\} \end{aligned}$$

Using Equations 5.32 and 5.33, we have

$$P_e = \frac{\sin(\pi/M)}{2\pi} \int_{-\pi/2}^{\pi/2} \frac{e^{-\lambda[1-\cos(\pi/M)\cos t]}}{1-\cos(\pi/M)\cos t} dt \quad (5.34)$$

For BPSK ( $M=2$ ), we have

$$P_e = \frac{1}{2} e^{-\lambda} \quad (5.35)$$

An approximate formula for Equation 5.34 is given as follows [15]:

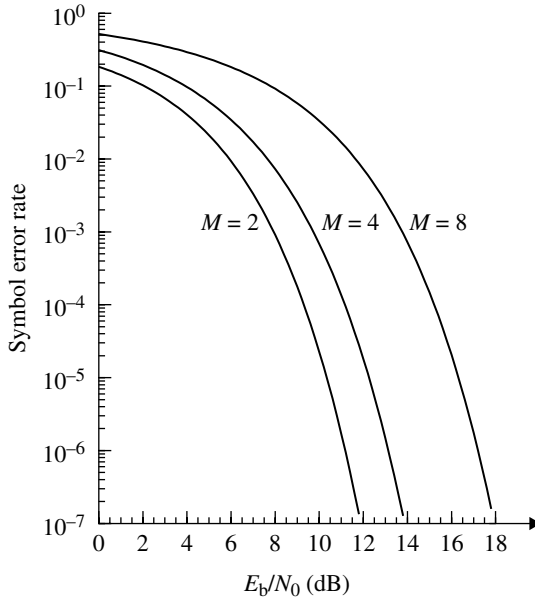
$$P_e \approx \sqrt{\frac{1+\cos(\pi/M)}{\cos(\pi/M)}} \text{erfc}\left(\left\{\lambda[1-\cos(\pi/M)]\right\}^{1/2}\right) \quad (5.36)$$

For Gray-coded QPSK, the bit error rate  $P_{eb}$  is given as

$$\begin{aligned} P_{eb} &= 2 \times \frac{1}{2} \text{Prob}\left(\frac{\pi}{4} < \psi < \frac{3}{4}\pi\right) + 2 \text{Prob}\left(\frac{3}{4}\pi < \psi < \pi\right) \\ &= \text{Prob}\left(\frac{\pi}{4} < \psi < \pi\right) + \text{Prob}\left(\frac{3}{4}\pi < \psi < \pi\right) \\ &= \frac{1}{4\sqrt{2}\pi} \int_{-\pi}^{\pi} \frac{e^{-\lambda[1-(\cos t)/\sqrt{2}]}}{1-(1/\sqrt{2})\cos t} dt + \frac{1}{4\sqrt{2}\pi} \int_{-\pi}^{\pi} \frac{e^{-\lambda[1+(\cos t)/\sqrt{2}]}}{1+(1/\sqrt{2})\cos t} dt \end{aligned} \quad (5.37)$$

Error rates for BPSK, QPSK, and 8PSK with differential detection are shown in Figure 5.26. We must remember that the above results are given with no inter-symbol interference and no correlation between noises at the sampling instant. These conditions are satisfied for a linear PSK system, where the square root of the Nyquist-I transfer characteristics is given for the transmit and receive bandpass filters and where the noise power spectrum is flat.

Satisfying the condition  $\omega_c T = 2n\pi$  is rather easy for the case of  $\omega_c \gg 1/T$ . In this case, if  $\omega_c T \neq 2n\pi$ , we can adjust the time delay  $T \rightarrow T + \Delta T$  to achieve  $\omega_c(T + \Delta T) = \omega_c T (1 + \Delta T/T) = 2n\pi$ . Since  $\omega_c \gg 1$ ,  $\Delta T/T$  can be set small. Small adjustments of time delay never cause significant problems in detecting the phase difference of two symbols.



**FIGURE 5.26** Symbol error rate for  $M$ -ary PSK with differential detection.

However, the phase error that is generated when  $\omega_c T \neq 2n\pi$  causes a significant degradation of error rate performance. The phase error is produced by deviation of the time delay or the carrier frequency. The symbol error rate when the phase error  $\Delta\theta$  is taken into consideration can be given as follows:

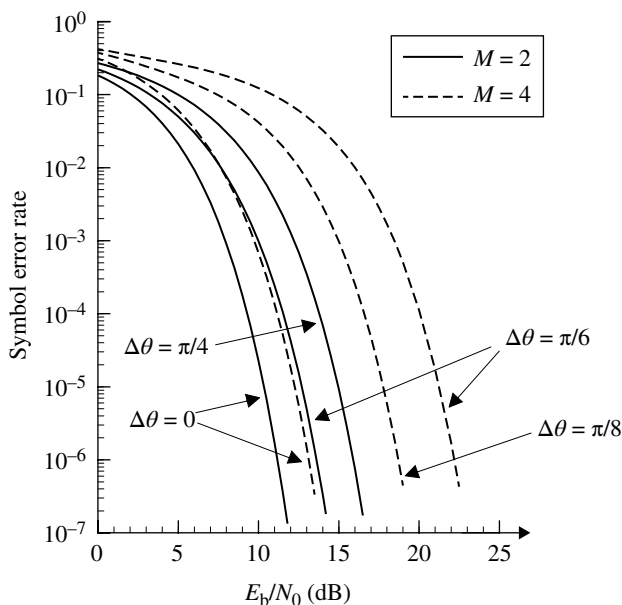
$$\begin{aligned}
 P_e(\Delta\theta) &= \text{Prob}\left(\frac{\pi}{M} - \Delta\theta < \psi < \pi\right) + \text{Prob}\left(-\pi < \psi < -\frac{\pi}{M} - \Delta\theta\right) \\
 &= \frac{\sin(\pi/M - \Delta\theta)}{4\pi} \int_{-\pi/2}^{\pi/2} \frac{e^{-\lambda[1 - \cos(\pi/M - \Delta\theta)\cos t]}}{1 - \cos(\pi/M - \Delta\theta)\cos t} dt \\
 &\quad + \frac{\sin(\pi/M + \Delta\theta)}{4\pi} \int_{-\pi/2}^{\pi/2} \frac{e^{-\lambda[1 - \cos(\pi/M + \Delta\theta)\cos t]}}{1 - \cos(\pi/M + \Delta\theta)\cos t} dt
 \end{aligned} \tag{5.38}$$

The symbol error rate performance for BPSK and QPSK is shown in Figure 5.27 with the phase error as a parameter.

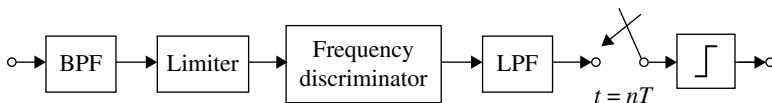
### 5.5.4 Frequency Discriminator Detection

A frequency discriminator detection system is shown in Figure 5.28. The input signal and noise are fed into a bandpass filter, followed by a limiter, a frequency discriminator, a low-pass filter, and a sample-and-decision circuit. This demodulator, except for the sample-and-decision circuit, is widely used for analog FM systems in mobile radio communication. Implementation of the limiter circuit is rather easy compared with an





**FIGURE 5.27** Effects of phase error on symbol error rate for BPSK and QPSK with differential detection.



**FIGURE 5.28** Frequency discriminator detection system.

automatic gain control circuit, especially for application to mobile radio communication, where there is fast fading and the received signal level has a large dynamic range.

The analysis of the error rate performance of the frequency discriminator detection system is generally difficult, since the demodulator is a highly nonlinear circuit. This is especially true when the effects of the bandpass filter and the postdetection filters on intersymbol interference must be considered at the same time. The digital FM signal can be expressed as

$$s(t) = A_0 s[\omega_c t + \varphi(t)]$$

where

$$\varphi(t) = k_F \int_{-\infty}^t \sum_{n=-\infty}^{\infty} a_n h(t-nT) dt = k_F \sum_{n=-\infty}^{\infty} a_n g(t-nT),$$

$a_n$  takes discrete values,  $h(t)$  is the impulse response of the premodulation filter, and  $g(t) = \int_{-\infty}^t h(t) dt$ .

The signal at the output of the bandpass filter is written as

$$s_{\text{BPF}}(t) = \text{Re} \left\{ A_0 [\cos \varphi(t) + j \sin \varphi(t)] * h_p(t) e^{j\omega_c t} \right\}$$

where  $h_p(t)$  is the low-pass equivalent impulse response of the IF bandpass filter. We assume that  $h_p(t)$  is real, that is,  $H_p(-\omega) = H_p^*(\omega)$ , where  $H_p(\omega) \leftrightarrow h_p(t)$ . Then  $s_{\text{BPF}}(t)$  is expressed as

$$s_{\text{BPF}}(t) = A(t) \cos[\omega_c t + \phi(t)]$$

where

$$\frac{A^2(t)}{A_0^2} = \{\cos \varphi(t) * h_p(t)\}^2 + \{\sin \varphi(t) * h_p(t)\}^2$$

$$\text{and } \phi(t) = \tan^{-1} \frac{\sin \varphi(t) * h_p(t)}{\cos \varphi(t) * h_p(t)}$$

The output signal of the bandpass limiter is expressed as

$$n_{\text{BPF}}(t) = n_x(t) \cos[\omega_c t + \phi(t)] - n_y(t) \sin[\omega_c t + \phi(t)] \quad (5.39)$$

where  $n_x(t)$  and  $n_y(t)$  are Gaussian random signals with zero mean and variance  $\sigma_n^2 = \frac{N_0}{2\pi} \int_{-\infty}^{\infty} |H_p(\omega)|^2 d\omega$ , and  $N_0/2$  is the input noise power spectrum density. The output signal of the bandpass limiter is expressed as

$$s_{\text{LIM}}(t) = \cos[\omega_c t + \phi(t) + \eta(t)]$$

where

$$\eta(t) = \tan^{-1} \frac{n_y(t)}{A(t) + n_x(t)}$$

Hence, the output signal of the frequency discriminator is given as

$$d(t) = \dot{\phi}(t) + \dot{\eta}(t)$$

where  $(\cdot)$  denotes the time differentiation.  $\dot{\phi}(t)$  corresponds to the information bearing signal. The noise term can be given as

$$\dot{\eta}(t) = \frac{[A(t) + n_x(t)]\dot{n}(t) - n_y(t)[\dot{A}(t) + \dot{n}(t)]}{[A(t) + n_x(t)]^2 + n_y^2(t)}$$

and is no longer Gaussian.

The output signal of post-demodulation filter with impulse response of  $h_d(t)$  becomes

$$d_{\text{LPF}}(t) = d(t) * h_d(t) = s_d(t) + n_d(t)$$

where  $s_d(t) \equiv \dot{\phi}(t) * h_d(t)$  and  $n_d(t) \equiv \dot{\eta}(t) * h_d(t)$ . A decision error occurs when  $d_{\text{LPF}}(nT)$  deviates over the threshold value(s). In order to get the error rate theoretically, we must know the statistical property of inter-symbol interference in signal  $s_d(t)$  and the probability density function of the noise  $n_d(t)$ . Since  $s_d(t)$  and  $n_d(t)$  are produced through nonlinear processes, it is impossible to find an expression for the error rate. The inter-symbol interference depends on the premodulation filter, modulation index, the bandpass filter, and the post-demodulation filter, as well as the data sequence.

The behavior of the noise term was investigated by S. O. Rice [16] using the concept of “clicks.”  $\dot{\eta}(t)$  is expressed with the continuous part and discontinuous or the click part as

$$\dot{\eta}(t) = \dot{\eta}_c(t) + \dot{\eta}_d(t) \quad (5.40)$$

the click part is expressed as

$$\dot{\eta}_d(t) = \sum_i 2\pi\delta(t - t_i) - \sum_j 2\pi\delta(t - t_j) \quad (5.41)$$

where the first and the second terms express positive and negative clicks, respectively. The click is an impulse that occurs randomly as shown schematically in Figure 5.29. The probabilities that clicks occur in a time  $T_0$  is assumed to obey to the Poisson distribution (Appendix 9.A):

$$P_N = \frac{(\lambda T_0)^N e^{-\lambda T_0}}{N!} \quad (5.42)$$

where  $N$  is the number of clicks and  $\lambda$  is the rate of occurrence of a click in an infinitely small time interval, or equivalently, average number of clicks in a unit time. The probability of occurrence of a click in a unit time is given [16] for a positive click as



**FIGURE 5.29** Schematic drawing of output noise of a frequency discriminator.

$$N_+(t) = \frac{r}{2} \left[ \left\{ 1 + \frac{f_i^2(t)}{r^2} \right\}^{1/2} \operatorname{erfc} \left\{ \rho(t) + \rho(t) \frac{f_i^2(t)}{r^2} \right\}^{1/2} - \frac{|f_i(t)|}{r} e^{-\rho(t)} \operatorname{erfc} \left\{ \frac{|f_i(t)|}{r} \sqrt{\rho(t)} \right\} \right] \quad (5.43)$$

and for a negative click as

$$N_-(t) = N_+(t) + |f_i(t)| e^{-\rho(t)} \quad (5.44)$$

where  $f_i(t) = \dot{\phi}(t) / 2\pi$  is the instantaneous frequency of the signal and  $f_i(t) \geq 0$  is assumed.  $\rho(t)$  is the signal-to-noise power ratio at the output of the bandpass filter and  $r$  is a parameter that is defined as

$$r = \frac{1}{2\pi} \frac{\langle \dot{n}_x^2(t) \rangle}{\langle n_x^2(t) \rangle} = \frac{1}{2\pi} \frac{\langle \dot{n}_y^2(t) \rangle}{\langle n_y^2(t) \rangle}$$

$\rho(t)$  is written as

$$\rho(t) = \frac{A^2(t)}{2\sigma_n^2}$$

and  $\sigma_n^2$  is rewritten as

$$\begin{aligned} \sigma_n^2 &= \frac{N_0}{2} \frac{1}{2\pi} \int_{-\infty}^{\infty} |H_p(\omega)| d\omega \\ &= \frac{N_0}{2} \int_{-\infty}^{\infty} |H_p(2\pi f)|^2 df \\ &= N_0 B_{\text{IF}} \left( f = \frac{\omega}{2\pi} \right) \end{aligned}$$

where  $B_{\text{IF}} = \int_0^{\infty} |H_p(2\pi f)|^2 df$  is the equivalent noise bandwidth of the bandpass filter.

It can be seen from Equation 5.44 that  $N_-(t) \geq N_+(t)$  ( $f \geq 0$ ). For  $f_i(t) < 0$ , it is understood that  $N_+(t)$  and  $N_-(t)$  correspond to the negative and positive clicks, respectively.

**5.5.4.1 Integrate-and-Dump Post-Demodulation Filter System** Error rate theories developed so far are different for the systems under consideration. One of them, for example, treats a system with a rectangular pulse shape (NRZ signaling) for transmit baseband signal and the integrate-and-dump filter as the postdetection filter as mentioned in Refs. [4] and [17–19]. In this system, no inter-symbol interference occurs, as long as the effect of the bandpass filter is ignored.

By integrating or equivalently filtering (Section 3.3.3) the output signal  $d(t)$  of the frequency discriminator over a time period  $(n-1)T \leq t \leq nT$ , we have

$$d_{\text{LPF}}(nT) = \phi(nT) - \phi(nT - T) + \eta(nT) - \eta(nT - T)$$

The signal term  $\phi(nT) - \phi(nT - T)$  is different from  $\varphi(nT) - \varphi(nT - T)$  due to inter-symbol interference, which depends on the data sequence. Let us denote

$$\Delta\phi(nT) \equiv \phi(nT) - \phi(nT - T) = \Delta\varphi(nT) + \delta\varphi(nT | \dots a_{n-1} a_n a_{n+1} \dots)$$

where  $\delta\varphi(nT | \dots a_{n-1} a_n a_{n+1} \dots)$  denotes the inter-symbol interference given the data sequence. From Equation 5.8, we have  $\Delta\varphi(nT) = \pi m a_n / a_{\text{max}}$ .

The noise term  $\Delta\eta(nT) \equiv \eta(nT) - \eta(nT - T)$  can be expressed from Equations 5.40 and 5.41 as

$$\Delta\eta(nT) = \eta_c(nT) - \eta_c(nT - T) + 2\pi[N^+(nT) - N^-(nT)]$$

where  $N^+(nT)$  and  $N^-(nT)$  are the number of positive and negative clicks occurred during a time period of  $(n-1)T \leq t \leq nT$ . The probability of  $N^+(nT)$  or  $N^-(nT)$  can be given from Equation 5.36 with substitution of  $N = N^+(nT)$  or  $N = N^-(nT)$ , and

$$\lambda T_0 = \int_{(n-1)T}^{nT} N_+(t) dt \quad \text{positive click}$$

or

$$= \int_{(n-1)T}^{nT} N_-(t) dt \quad \text{negative click}$$

When  $\eta_c(nT)$ ,  $\eta_c(nT - T)$ , and  $N^+(nT)$ ,  $N^-(nT)$  are independent from each other, the probability density function of  $\Delta\eta(nT)$  is given by the convolution of the probabilities for those variables.

In Ref. [18], the error rate is given for 2-level FM by numerical calculation of the inter-symbol interference  $\delta\varphi(nT | \dots a_{n-1} a_n a_{n+1} \dots)$  with a random data sequence of 30-bits. In Refs. [19] and [20], a closed form formula of bit error rate for narrow-band system is obtained by assuming a data sequence of 3 bits. The author's result is cited with his notation and is as follows:

$$P_e = P_{\text{continuous}} + P_{\text{click}}$$

where

$$\begin{aligned} P_{\text{click}} &= \frac{h}{4} e^{-R_d} + \int_0^\pi \frac{d}{dx} \left\{ \tan^{-1} \frac{-m \cos x}{1 - n \cos(2x + \delta)} \right\} \\ &\exp \left\{ -R_a \frac{[1 - n \cos(2x + \delta)]^2 + m^2 \cos^2 x}{(1 - n \cos \delta)^2 + m^2} \right\} \frac{1}{4\pi} dx \\ P_{\text{continuous}} &= \frac{1}{4} [P\{\psi > \Delta\phi | 111\} + P\{\psi > \Delta\psi | 010\} + 2P\{\psi > \Delta\phi | 011\}] \end{aligned}$$

111, 010, 011 are bit patterns

$$P\{\psi > \Delta\phi\} = \int_{\Delta\phi}^{\pi} p(\psi) d\psi$$

where  $p(\psi) =$

$$\frac{e^{-U}}{2\pi} \left[ \cosh V + \frac{1}{2} \int_0^{\pi} d\alpha (U \sin \alpha + W \cos \psi) \cdot \cosh(V \cos \alpha) \cdot e^W \sin \alpha \cos \psi \right]$$

Parameters  $\Delta\phi$ ,  $U$ , and  $V$  are given for the bit patterns as follows:

1. “111” Bit Pattern

$$\Delta\phi = \pi h, \quad U = R_d, \quad V = 0$$

2. “010” Bit Pattern

$$\Delta\phi = 2 \tan^{-1} \frac{m}{1 - n \cos \delta}, \quad U = R_a, \quad V = 0$$

3. “011” Bit Pattern

$$\Delta\phi = \frac{\pi h}{2} + \tan^{-1} \frac{m}{1 - n \cos \delta}, \quad U = \frac{R_a + R_d}{2}, \quad V = \frac{R_a - R_d}{2}$$

and  $h = 2f_d T, f_1 = 1/2T$ ,

$$W = (U^2 - V^2)^{1/2}, \quad m = \frac{2h^2 |H(f_1)|}{1 - h^2} \cot \frac{\pi h}{2}, \quad n = \frac{2h^2 |H(2f_1)|}{4 - h^2},$$

$$\delta = \angle H(2f_1) - 2\angle H(f_1),$$

$$R_a = \frac{E_b}{N_0} \left\{ \frac{\sin^2(\pi h / 2)}{\pi h / 2} \right\} \frac{(1 - n \cos \delta)^2 + m^2}{T \int_{-\infty}^{\infty} |H(f)|^2 df}, \quad R_d = \frac{E_b}{N_0} \frac{|H(f_d)|^2}{T \int_{-\infty}^{\infty} |H(f)|^2 df}$$

where  $H(f)$  is the transfer function of the bandpass filter,  $N_0$  is the *one-sided* spectral density of the input noise,  $f_d$  is the frequency deviation due to modulation,  $T$  is the bit duration, and  $h$  is the modulation index (reader should not confuse the notations used here and in other chapters of this book).

**5.5.4.2 General Post-Demodulation Filter System** A pulse shape with a narrower spectrum than that of the rectangular pulse shape is desired for the modulating signal. This is because we can get a narrower spectrum for modulated signal as described in Section 5.3. In this case, the integrate-and-dump filter is not appropriate for the post-demodulation filter. Schilling et al. [21] treated the system with a Gaussian post-demodulation filter. Following this, the Nyquist- and partial response-filtered multilevel FM systems are analyzed as follows [22].

Let us assume that the error rate is given as

$$P_e = P_g + P_c = \frac{1}{N} \sum_{n=1}^N \{P_g(nT) + P_c(nT)\}$$

where  $P_g$  is the error rate with Gaussian noise and  $P_c$  is the error rate with click noise. (This phenomenological assumption is supported by computer simulation experiments.)  $P_g(nT)$  and  $P_c(nT)$  denote error rates at the symbol time  $t=nT$ , due to the Gaussian and click noises, respectively.  $N$  is the length of the data sequence under consideration.

**Error Rates Due to Gaussian Noise** In order to account for the effect of the modulated signal on demodulated noise, we express the noise signal at the bandpass filter in a way different from Equation 5.39 as follows:

$$n_{\text{BPF}}(t) = n_x(t) \cos(\omega_c t) - n_y(t) \sin(\omega_c t)$$

The band-pass limited signal is then

$$n_{\text{LIM}}(t) = \cos(\omega_c t + \phi(t) + \psi(t))$$

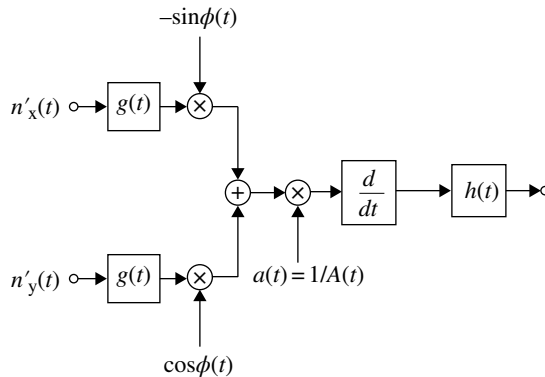
where

$$\psi(t) = \tan^{-1} \frac{n_y(t) \cos \phi(t) - n_x(t) \sin \phi(t)}{A(t) + n_x(t) \cos \phi(t) + n_y(t) \sin \phi(t)}$$

When the signal-to-noise power ratio is high, using  $|A(t)| \gg |n_x(t)|, |n_y(t)|$ ,  $\psi(t)$  is linearly approximated as

$$\psi(t) \approx \frac{n_y(t) \cos \phi(t) - n_x(t) \sin \phi(t)}{A(t)} \quad (5.45)$$

Hence  $\psi(t)$  and also the discriminator output  $\frac{d}{dt} \psi(t)$  become a Gaussian variable. The equivalent circuit for the approximated noise is represented in Figure 5.30.



**FIGURE 5.30** Equivalent circuit for noise at high signal-to-noise power ratio.

The expected noise power at the output of the post-demodulation filter becomes as follows (Appendix 5.B):

$$\langle N_g(t) \rangle = N_0 \int_{-\infty}^{\infty} |N(t, \tau)|^2 d\tau$$

where

$$|N(t, \tau)|^2 = \{h_d(t) * [g(t - \tau) a(t) \cos \phi(t)]\}^2 + \{h_d(t) * [g(t - \tau) a(t) \sin \phi(t)]\}^2$$

$g(t)$  is the low-pass equivalent impulse response of the bandpass filter,  $h_d(t)$  is the impulse response of the differentiating circuit including the post-demodulation filter, and  $a(t) \equiv 1/A(t)$ .

Now the error rate due to the Gaussian noise at time slot  $t = nT$  is given as

$$P_g(nT_s) = Q\left(\frac{d_n^+}{\sigma_n}\right) + Q\left(\frac{d_n^-}{\sigma_n}\right) \quad (5.46)$$

where

$$d_n^\pm = 2\pi\{f_m \pm \Delta f_m(nT)\}, \quad \sigma_n = \left[\langle N_g(t) \rangle\right]^{1/2} \quad (5.47)$$

$$Q(y) = \frac{1}{\sqrt{2\pi}} \int_y^\infty \exp\left(-\frac{x^2}{2}\right) dx$$

and  $2\pi f_m$  is half the signal distance at the output of the postdetection filter. The term  $2\pi \Delta f_m(nT)$  represents inter-symbol interference at the sampling instant  $t = nT$ . These values can be obtained by numerical calculation assuming a data sequence.  $\sigma_n^2$  is the expected noise power at the sampling instant. When the data takes the highest (lowest) value, the first (second) term in the right-hand side of Equation 5.46 may be ignored.



**Error Rates due to Click Noise** A positive click (impulsive) noise that occurs at  $t=t_1$  generates a waveform  $2\pi h(t-t_1)$  at the output of the postdetection low-pass filter;  $h(t)$  is the impulse response of the filter. An error occurs if the following relations hold (Fig. 5.31)

$$2\pi h(nT-t_1) (\geq 0) \geq 2\pi f_m - 2\pi \Delta f_m(nT) \quad (5.48a)$$

$$2\pi h(nT-t_1) (< 0) \leq -[2\pi f_m + 2\pi \Delta f_m(nT)] \quad (5.48b)$$

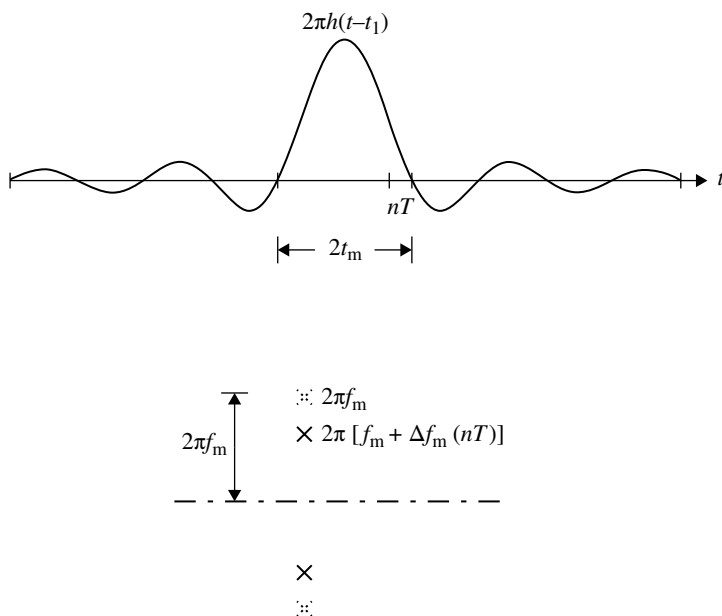
For simplicity of discussion, we only consider the time period  $(2t_m)$  where  $h(t) \geq 0$ , then Equations 5.48a and 5.48b correspond to the positive and negative clicks, respectively. Furthermore, we neglect the possibility that two or more clicks cause an error, since the probability of that is low. (This cannot be valid for a wideband digital FM system with signal-to-noise power ratio under the threshold value.) Hence, the error rate due to the click noise is given as

$$P_c(nT) = \int_{D_{n1}^+}^{D_{n2}^+} N^+(t-nT) dt + \int_{D_{n1}^-}^{D_{n2}^-} N^-(t-nT) dt$$

where  $D_{n1}^\pm$  and  $D_{n2}^\pm$  are parameters that satisfy the following equations

$$h(D_{n1}^+) = h(D_{n2}^+) = f_m - \Delta f_m(nT) \geq 0 \quad (5.49a)$$

$$h(D_{n1}^-) = h(D_{n2}^-) = f_m + \Delta f_m(nT) \geq 0 \quad (5.49b)$$



**FIGURE 5.31** Click noise at the output of post-demodulation filter.

For an  $h(t)$  that is symmetrical with respect to time, we have  $D_{n1}^+ = D_{n2}^+$  and  $D_{n1}^- = D_{n2}^-$ . No positive (negative) clicks cause errors at the symbol that takes the highest (lowest) level.

If the bandwidth  $B_{\text{IF}}$  of the bandpass filter is wide, we can ignore the intersymbol interference due to the filter and the effect that modulation has on the Gaussian noise. This condition can be supported by experiments if the bandwidth is wider than the Carson bandwidth:

$$B_{\text{IF}} \geq 2(\Delta F_m + f_b)$$

where  $\Delta F_m$  is the maximum frequency deviation and  $f_b$  is the highest frequency of the modulating baseband signal.

Furthermore, we assume that the baseband channel, including the premodulation filter and the postdetection filter, never produces intersymbol interference, except for the partial response system. The output signal of the post-demodulation filter is expressed as follows:

$$s_0(t) = 2\pi f_m a_i h_T(t) * h_R(t) \quad ((i-1)T \leq t \leq iT)$$

where the symbol data  $a_i$  takes discrete values such that there is a signal distance of  $2d$  between the neighboring values, and  $h_T(t)$  and  $h_R(t)$  are the impulse response of the premodulation and post-demodulation filter, respectively. It is understood that

$$\int_{-\infty}^{\infty} h_T(t) dt = H_T(\omega = 0) = T$$

where  $H_T(\omega) \leftrightarrow h_T(t)$ . For the sake of simplicity, we let

$$h_T(t) * h_R(t) \big|_{t=iT} = 1 \quad (5.50)$$

The signal distance  $2d$  becomes

$$2d = 4\pi f_m$$

The transfer function for the baseband signal is

$$H(\omega) = H_T(\omega) H_R(\omega) = k_0 H_I(\omega)$$

where  $H_R(\omega) \leftrightarrow h_R(t)$ ,  $k_0$  is a constant and  $H_I(\omega)$  meets the Nyquist-I criterion.

For  $H_I(\omega)$  with  $H_I(0) = 1$ , we have  $h_I(0) = 1/T$  (Eq. 3.9), and we let

$$H_T(\omega) = TH_{IT}(\omega),$$

$$H_R(\omega) = H_{IR}(\omega),$$

where

$$\begin{aligned} H_{IT}(\omega = 0) &= H_{IR}(\omega = 0) = 1, \\ H_{IT}(\omega) H_{IR}(\omega) &= H_I(\omega). \end{aligned}$$

then the condition given by Equation 5.50 is satisfied.

When the bandwidth of the bandpass filter is so wide that the interference due to it is ignored, the amplitude of the output signal of the bandpass filter  $A(t)$  becomes a constant value,  $A$ . For a high signal to noise power ratio, the output signal of the frequency discriminator becomes

$$n_d(t) \approx \frac{d}{dt} \psi(t)$$

where

$$\psi(t) = \frac{n_y(t) \cos \varphi(t) - n_x(t) \sin \varphi(t)}{A}$$

The autocorrelation function of  $\psi(t)$  becomes

$$R_\psi(\tau) \equiv \langle \psi(t) \psi(t + \tau) \rangle = \frac{1}{A^2} R_{n_x}(\tau) \text{Re}\{R_\varphi(\tau)\} \quad (5.51)$$

where

$$R_{n_x}(\tau) = \langle n_x(t) n_x(t + \tau) \rangle = \langle n_y(t) n_y(t + \tau) \rangle$$

and

$$R_\varphi(\tau) = \langle e^{j\varphi(t)} e^{-j\varphi(t+\tau)} \rangle$$

Taking the Fourier transform of Equation 5.51, we have the PSD of  $\psi(t)$ : the convolution integral of the PSD of  $n_x(t)$  and FM signal  $e^{j\varphi(t)}$ . When we assume a wideband bandpass filter, the effect of FM on the noise PSD can be neglected. Hence the power spectrum density of  $n_d(t)$  is given as follows:

$$S_{n_d}(\omega) = \frac{N_0 \omega^2}{A^2} \quad (-\pi B_{\text{IF}} < \omega < \pi B_{\text{IF}})$$

where  $N_0/2$  is the input noise power density and  $B_{\text{IF}}$  is the bandpass filter bandwidth. The average noise power at the output of the post-demodulation filter is as follows:

$$\langle N_s \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{N_0 \omega^2 |H_R(\omega)|^2}{A^2} d\omega \quad (5.52)$$

We define a normalized noise bandwidth by

$$W_{\text{eq}} = \frac{\int_{-\infty}^{\infty} \omega^2 |H_R(\omega)|^2 d\omega}{\int_{-\omega_s/2}^{\omega_s/2} \omega^2 d\omega \left( = \frac{2}{3} \left( \frac{\omega_s}{2} \right)^3 \right)}$$

where  $\omega_s = 2\pi/T$ .

Rewriting Equation 5.52

$$\langle N_g \rangle = \frac{1}{24} \frac{N_0 f_s}{A^2 / 2} (2\pi f_s)^2 W_{\text{eq}} = \frac{1}{24} \frac{1}{E_s / N_0} (2\pi f_s)^2 W_{\text{eq}}$$

Rewriting  $d_n^\pm = d_0 = 2\pi f_m$ ,  $\sigma_n = \sigma_0 = [\langle N_g \rangle]^{1/2}$  in Equations 5.46 and 5.47 and taking into consideration error probability  $\text{Prob}(a_m)$  for the highest or lowest levels, we have

$$P_g = 2[1 - \text{Prob}(a_m)] Q\left(\frac{d_0}{\sigma_0}\right) \quad (5.53)$$

where

$$\frac{d_0}{\sigma_0} = \frac{2f_m}{f_s} \sqrt{\frac{6}{W_{\text{eq}}} \frac{E_s}{N_0}} = \sqrt{\frac{6}{W_{\text{eq}}} \frac{E_s}{N_0}} m \quad (5.54)$$

where  $m$  is modulation index. Considering the Nyquist-I, duobinary, and class-II partial response digital FM systems with their baseband filter characteristics (Eqs. 3.8, 3.17 and 3.19) equally split between the premodulation and post-demodulation filter, and normalizing them to get  $H_R(\omega = 0)$ , we have the normalized noise bandwidth and is as follows:

$$W_{\text{eq}} = \begin{cases} 1 + 3\left(1 - \frac{8}{\pi^2}\right)\alpha^2 & (\text{Nyquist - I}) \\ 6\left(\frac{1}{\pi} - \frac{8}{\pi^3}\right) & (\text{duobinary}) \\ \frac{1}{2} - \frac{3}{\pi^2} & (\text{class - II partial response}) \end{cases} \quad (5.55)$$

where  $\alpha$  is the roll-off factor.

For the integrate-and-dump filter, the integral diverges and we could not define the normalized bandwidth. For this system, the noise power must be calculated taking into consideration the bandpass filter:

$$\langle N_g \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{N_0 \omega^2 |G(\omega) H_R(\omega)|^2 d\omega}{A^2}$$

where  $G(\omega)$  is the low-pass equivalent transfer function of the band-pass filter. From this fact, we can see that the noise power is higher for the integrate-and-dump filter system than for other systems when the band-pass filter bandwidth is much wider than the postdetection filter bandwidth. This is usually true for multilevel FM or partial response systems.

*Error Rates due to Click Noise* Assuming  $\Delta f_m(nT) = 0$ ,  $D_{n1}^+ = D_{n1}^- = D_{n1}$  and  $D_{n1}^+ = D_{n1}^- = D_{n2}$  in Equations 5.49a and 5.49b, we have the error rate due to click noise as follows:

$$P_c = \frac{1}{N} \sum_{n=1}^N \left\{ \int_{D_{n1}}^{D_{n2}} [N_+(t - nT) + N_-(t - nT)] dt \right\}$$

The error rates are still dependent on the modulating signal  $f_i(t)$  in Equation 5.43. If we assume dc demodulating signals, that is,  $f_i(t) = a_n f_m$ , then  $\rho(t)$  becomes a constant value and  $N_+(a_n f_m)$  and  $N_-(a_n f_m)$  denote  $N_+(t)$  and  $N_-(t)$  in Equations 5.43 and 5.44. In this case, we have

$$P_c = \sum_{a_n} \text{Prob}(a_n) (D_{n2} - D_{n1}) [N_+(a_n f_m) + N_-(a_n f_m)] \quad (5.56)$$

*The Optimum Modulation Index* In an FM system with discriminator detection, the wideband gain is well known: the demodulated signal to noise power ratio increases with the increase of the maximum frequency deviation or the modulation index (see for example Equation 5.54). Figure 5.32 shows error rates for a 2-level FM as a function of the modulation index. The error rates are calculated using Equations 5.53, 5.55, and 5.56 without intersymbol interference due to a bandpass filter. The bandpass filter bandwidth  $B_{\text{IF}}$  is chosen as the Carson bandwidth, that is,  $B_{\text{IF}} = (m + 1 + \alpha) f_b$ , where  $m$  is the modulation index,  $\alpha$  the roll-off factor, and  $f_b$  the bit rate frequency. Therefore, the carrier to noise power ratio at the output of the bandpass filter becomes  $C/N = E_b f_b / N_0 B_{\text{IF}} = E_b / N_0 / (m + 1 + \alpha)$ . When we increase the modulation index from a low initial value, the error rates first decrease due to the decrease in the Gaussian noise. After reaching the minimum value, it increases due to the increase in the click noise (threshold effect). The optimum modulation index for the 2-level FM at a bit error rate of  $10^{-2} \sim 10^{-3}$  is around 0.5. If we consider a lower error rate and hence a higher  $E_b/N_0$ , the optimum modulation index takes a higher value.

In the above argument, we made use of the flat transfer characteristics of the bandpass filter. For a high modulation index, the 2-level FM spectrum shows two peaks at the frequencies of  $f_c \pm \Delta f_m$ , where  $\Delta f_m$  is the maximum deviation (see Fig. 5.10). In this case, we can use a bandpass filter whose transfer characteristic is matched to the spectrum and then the threshold effect is softened.

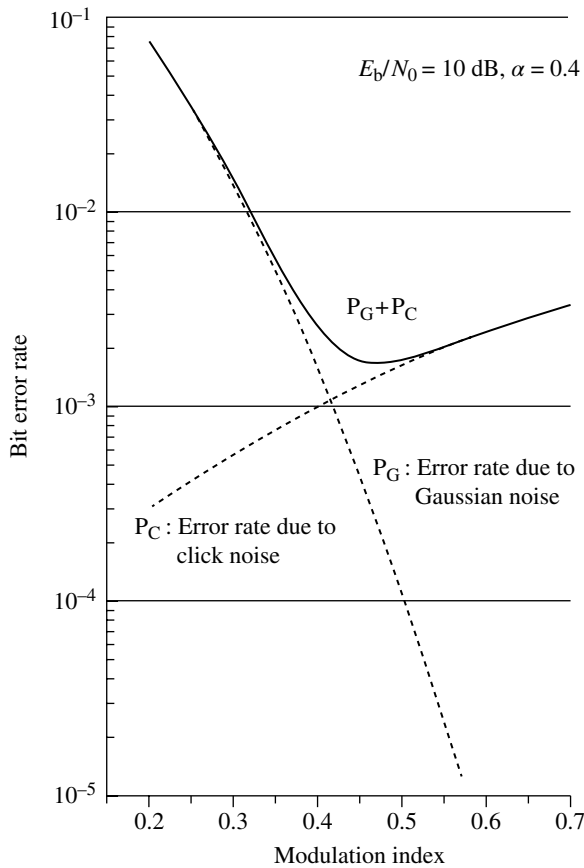


FIGURE 5.32 Error rates for 2-level FM as a function of modulation index.

### 5.5.5 Error Rates in Fading Channels

In a multipath fading channel, for example, mobile radio channel, the signal is subjected to random fluctuations in power levels, phases, and the Doppler frequency shifts. Furthermore, intersymbol interference appears when the signal bandwidth becomes comparable to the coherence bandwidth of the channel (Section 4.4). These phenomena degrade performance.

**5.5.5.1 Error Rates due to Level Fluctuation** The average error rate is given as

$$\langle P_e \rangle = \int_0^{\infty} P_e(\gamma) p(\gamma) d\gamma \quad (5.57)$$

where  $\gamma$  is the signal to noise power ratio  $S/N$  or symbol energy to noise power density ratio  $E_s/N_0$ ,  $P_e(\gamma)$  is error rate for a given  $\gamma$ , and  $p(\gamma)$  is the probability density of  $\gamma$ . If we denote the constant signal envelope by  $u$ , we have

$$\gamma \equiv \frac{E_s}{N_0} = \frac{u^2 T}{2N_0} \quad (5.58)$$

where  $T$  is the symbol duration.

Consider the Rayleigh fading discussed in Section 4.3, from Equation 4.11 with  $b = u_0^2 / 2$ , we have

$$p(u) = \frac{2u}{u_0^2} e^{-\frac{u^2}{u_0^2}} \quad (0 \leq u \leq \infty) \quad (5.59)$$

where  $u_0^2 = \langle u^2 \rangle$ .

From Equations 5.58 and 5.59, we have

$$p(\gamma) = \frac{1}{\gamma_0} e^{-\gamma/\gamma_0} \quad (0 \leq \gamma \leq \infty) \quad (5.60)$$

For a Rayleigh distribution signal envelope  $u$ , the transformation of the variable  $\gamma = ku^2$  does not depend on the coefficient  $k$ . Therefore the distribution density function  $p(\gamma)$  also can be applied to  $E_b / N_0 = (E_s / N_0)(T_b / T)$ , where  $T_b$  is the bit duration.

For BPSK and Gray coded QPSK with coherent demodulation, the bit error rate  $P_e(\gamma)$  is given as  $P_e(\gamma) = (1/2) \operatorname{erfc}(\sqrt{E_b / N_0})$  (Eqs. 5.18a and 5.21). Inserting  $P_e(\gamma)$  into Equation 5.57 and using Equation 5.60, we have the average error rates

$$\begin{aligned} \langle P_e \rangle &= \int_0^\infty \frac{1}{2} \operatorname{erfc}(\sqrt{\gamma}) \cdot \frac{1}{\gamma_0} e^{-\gamma/\gamma_0} d\gamma = \frac{1}{2} \left\{ 1 - \frac{1}{\sqrt{1 + 1/\gamma_0}} \right\} \\ &\approx \frac{1}{4\gamma_0} (\gamma_0 \gg 1) \end{aligned}$$

For ASK and FSK with noncoherent detection, and BPSK with differential detection, we can express  $P_e(\gamma)$  as (Eqs. 5.30, 5.31, and 5.35):

$$P_e(\gamma) = \frac{1}{2} e^{-\alpha\gamma}$$

where

$$\alpha = \begin{cases} 1/2 & \text{ASK and FSK with noncoherent detection} \\ 1 & \text{BPSK with differential detection} \end{cases}$$

The average error rate is given by

$$\langle P_e \rangle = \int_0^\infty \frac{1}{2} e^{-\alpha\gamma} \frac{1}{\gamma_0} e^{-\gamma/\gamma_0} d\gamma$$

$$= \frac{1}{2} \frac{1}{1 + \alpha\gamma_0} \quad (5.61a)$$

$$\approx \frac{1}{2\alpha\gamma_0} \quad (\alpha\gamma_0 \gg 1) \quad (5.61b)$$

**5.5.5.2 Error Rates due to the Random FM Effect** Received signals experience random fluctuations in phase or frequency (random FM effect), as well as in envelope in fading channels. The error due to random FM effects could not be reduced by increasing the signal power and it is called “irreducible error.” The analysis of irreducible error rates for coherent detection is difficult, since it includes a carrier recovery system, which is a highly nonlinear circuit. Here frequency discriminator detection and differential detection are considered.

For discriminator detection, an error occurs when the instantaneous frequency deviation due to the random FM exceeds a threshold value at the sampling instant. The probability density function of the instantaneous frequency  $\dot{\theta}$  is given by Equation 4.20. For a vertically polarized signal, we have

$$p(\dot{\theta}) = \frac{(\pi f_m)^2}{[\dot{\theta}^2 + 2(\pi f_m)^2]^{3/2}}$$

where  $f_m$  is the maximum Doppler frequency. The probability that  $\dot{\theta}$  exceeds a threshold value  $\Delta\omega_d$  becomes

$$\begin{aligned} \langle P \rangle &= \text{Prob}(\dot{\theta} > \Delta\omega_d) = \int_{\Delta\omega_d}^{\infty} \frac{(\pi f_m)^2}{[\dot{\theta}^2 + 2(\pi f_m)^2]^{3/2}} d\dot{\theta} \\ &= \frac{1}{2} \left[ 1 - \frac{1}{\sqrt{1 + 2^{-1} (f_m / \Delta f_d)^2}} \right] \end{aligned} \quad (5.62a)$$

$$\approx \frac{1}{8} \left( \frac{f_m}{\Delta f_d} \right)^2 \quad (f_m \ll \Delta f_d) \quad (5.62b)$$

where  $\Delta f_d = \Delta\omega_d / 2\pi$  is one-half of the frequency separation between signals. Considering the error in the highest (lowest) level, the error rate due to the random FM becomes

$$\langle P_e \rangle = 2 \left\{ 1 - \frac{1}{2} P(b_M) \right\} \langle P \rangle \quad (5.63)$$

where  $P(b_M)$  denotes the probability that the signal takes the highest or lowest level.



From Equation 5.62b, we can see that the frequency separation  $\Delta f_d$  must be large enough compared with the maximum Doppler frequency  $f_m$ . This means that low speed data transmission requires a high modulation index.

For BPSK with bandpass matched filter and differential detection, the error rate for Rayleigh fading is given as (Eqs. 4.2–4.7) in [23]:

$$\langle P_e \rangle = \frac{1 + \lambda_0 [1 - J_0(2\pi f_m T)]}{2(1 + \lambda_0)}$$

where  $\lambda_0$  is the average energy per bit to noise power density,  $J_0(\cdot)$  is the zeroth-order Bessel function of the first kind, and  $T$  is the bit duration. For a quasi-static fading ( $f_m \rightarrow 0$ ),  $\langle P_e \rangle$  becomes Equation 5.61a ( $\alpha = 1$ ). For  $\lambda_0 \rightarrow \infty$ ,  $\langle P_e \rangle$  corresponds to the irreducible error rate. It becomes

$$\langle P_e \rangle = \frac{1 - J_0(2\pi f_m T)}{2} \approx \frac{1}{2} (\pi f_m T)^2 \quad (f_m T \ll 1)$$

where approximation  $J_0(x) \approx 1 - (x/2)^2$  ( $x \ll 1$ ) is used.

**5.5.5.3 Error Rates due to Frequency-Selective Fading** As the signal bandwidth becomes as wide as the coherence bandwidth of the channel, the signal experiences frequency selective fading; the transfer function of the channel is not flat over the signal bandwidth, causing intersymbol interference. The error rate due to the frequency selective fading depends on the pulse waveform, demodulation system, as well as the statistics of the fading. Therefore it is difficult to obtain general formulas for error rates due to frequency selective fading. The frequency correlation function of the channel is defined as

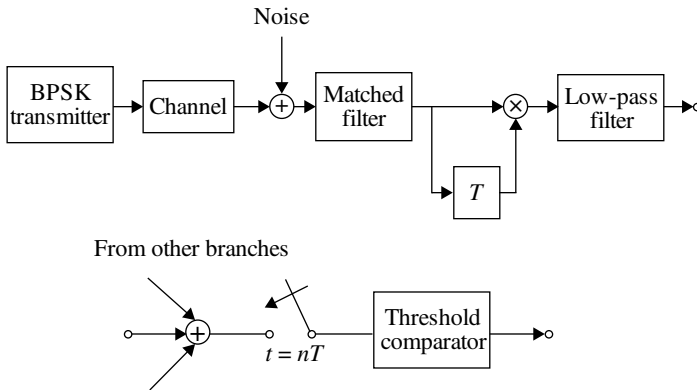
$$R(f) = \langle H(f_0) H^*(f_0 + f) \rangle$$

where  $H(f)$  is the channel transfer function and  $\langle \cdot \rangle$  means ensemble average.

Bello and Nelin [24] studied the effect of the frequency selective fading on the error rates of incoherent and differentially coherent (delay detection) matched filter demodulation of binary FSK and BPSK, including postdetection diversity. They assumed that the channel is time-invariant (quasi-stationary) and its transfer function is a sample function from a stationary complex valued zero mean Gaussian random process. Also the intersymbol interference depends on the data patterns. In their analysis, the intersymbol interference from only adjacent bits is considered.

Based on their analysis, Bailey and Lindenlaub [25] discussed the matched filtered differentially coherent BPSK receiver with postdetection diversity (Fig. 5.33) for a few pulse waveforms and frequency correlation functions. The pulse waveforms were an NRZ pulse and raised cosine spectrum signal (Eq. 3.8 with  $\alpha = 1$ ). The considered frequency correlation functions are the Gaussian-shaped (G-F channel) as

$$R(f) = 2\sigma^2 \exp\left(-\frac{4f^2}{B_c^2}\right)$$



**FIGURE 5.33** Matched filtered differentially coherent BPSK receiver with postdetection diversity.

and the sinc-function type (S-F channel) as

$$R(f) = 2R_0T_m \sin c(2fT_m)$$

where  $\text{sinc}(x) = \sin(\pi x)/(\pi x)$ .

They obtained compact expression for irreducible error rates with the NRZ pulse waveform as

$$P_e \approx \frac{1}{4} \binom{2L-1}{L} \left[ (2c_2)^L + 2(c_2 - c_1^2)^L \right] d^{2L}$$

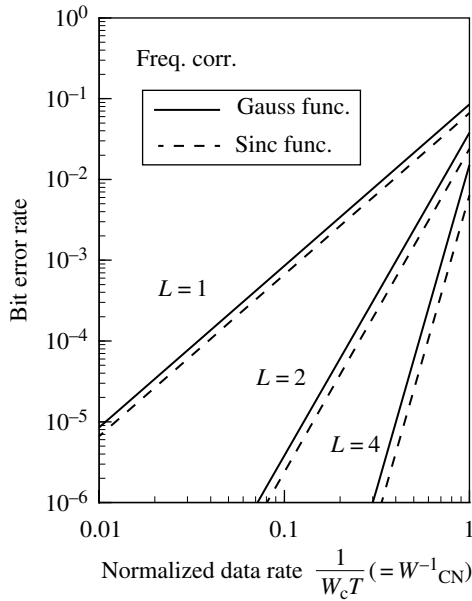
where  $L$  is the number of diversity branches and

$$d = \begin{cases} \frac{1}{TB_c} & (\text{G-F channel}) \\ \frac{T_m}{T} & (\text{S-F channel}) \end{cases}$$

$$c_1 = \begin{cases} \frac{1}{\pi\sqrt{\pi}} & (\text{G-F channel}) \\ \frac{1}{4} & (\text{S-F channel}) \end{cases}$$

$$c_2 = \begin{cases} \frac{1}{\pi^2} & (\text{G-F channel}) \\ \frac{1}{6} & (\text{S-F channel}) \end{cases}$$

The parameter  $d$  is the relative data rate: it is proportional to the ratio of the data rate  $1/T$  to the channel coherence bandwidth. The coherent bandwidth can be defined by the frequency range where the correlation function decreases to  $1/e$ . Then the



**FIGURE 5.34** Irreducible error rates for differential detection of BPSK with diversity reception.

normalized coherence bandwidth  $W_{\text{CN}}$ , the coherence bandwidth normalized by data rate, becomes  $W_{\text{CN}} = d^{-1}$  for the G-F channel and  $W_{\text{CN}} = 0.7 d^{-1}$  for the S-F channel. The irreducible error rates are shown as a function of  $W_{\text{CN}}$  in Fig. 5.34 for  $L=1$  (no diversity),  $L=2$ , and  $L=4$ .

The error rates due to the frequency selective fading can be reduced by introducing diversity reception when data rate is relatively small.

**5.5.5.4 Error Rates due to Co-channel Interference** Evaluating the effect of co-channel interference on error rates becomes important for estimation of co-channel reuse distance and hence the spectrum efficiency in a cellular system. Nevertheless, it has not been well analyzed theoretically. Hirade et al. [26, 27] discussed the error rates of digital FM with frequency discriminator and differential detection, taking the co-channel interference into consideration for fast Rayleigh fading channels. Their results of the irreducible error rates due to co-channel interference under the condition of quasi-static Rayleigh fading are given as

$$P_e = \frac{1}{2} \left[ 1 - \frac{\Lambda \sin(m\pi)}{\left\{ (\Lambda + 1)^2 - \left[ \Lambda \cos(m\pi) + \frac{1}{2} \left( \cos(m\pi) + \frac{\sin(m\pi)}{m\pi} \right) \right]^2 \right\}^{1/2}} \right] \quad (5.64a)$$

differential detection

$$P_e = \frac{1}{2(\Lambda + 1)} \quad \text{discriminator detection} \quad (5.64b)$$

where  $\Lambda$  is the carrier to co-channel interference signal power ratio ( $C/I$ ).

For  $m=0.5$  (minimum shift keying) and  $\Lambda \gg 1$ , Equation 5.64a reduces to Equation 5.64b. If we compare this result with the error rate (Eq. 5.61a) due to noise in quasi-static Rayleigh fading, the same error rate is given for the same value of  $C/I$  and  $C/N$ . Usually values of the error rates for Gaussian noise with the same value of  $C/N$  and  $C/I$  is a safe (i.e., higher) estimation for relatively low error rates. This is because the probability that the interference signal level takes a high value is less than that of the Gaussian noise.

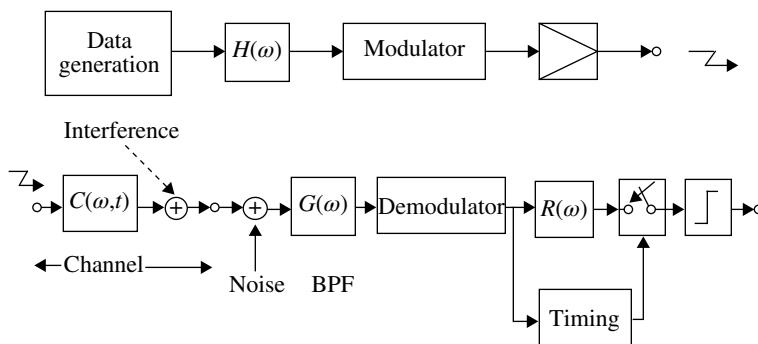
**5.5.5.5 Error Rates for QPSK in Fading Channels with Diversity Reception** The theoretical analyses in Refs. [28] and [29] are valuable. They found exact expressions for the error rates for Rayleigh fading channels with the diversity reception taking into account the effects of additive noise, random FM, co-channel interference, and frequency selective fading, simultaneously.

## 5.6 COMPUTER SIMULATION OF TRANSMISSION SYSTEMS

Main performance metrics of a digital transmission system are power spectrum of the transmitted signal, relative adjacent channel interference power, error rates, eye diagram, and co-channel interference. Those performance metrics must be evaluated for different modulation/demodulation schemes, combinations of transmit and receive filters, effects of carrier frequency offset, nonlinear distortion of a power amplifier, fading channel, diversity reception system, etc. Some of these metrics can be estimated through theoretical analysis; otherwise they must be evaluated by experiments. Implementing an experimental system is time-consuming especially for finding optimal parameters for the system. Furthermore, it is sometimes difficult for unskilled engineers.

Computer simulation experiments can be a powerful method for evaluating and designing digital modulation/demodulation systems, as well as other parts of a digital communication system. Operation of the experimental system is simulated by software programs on a computer. For example, Figure 5.35 shows an experimental system to be simulated.

For the digital test data, a pseudo-noise data sequence that has a flat power spectrum is desirable, since it generates different data patterns. For this purpose, the  $M$ -sequence is widely used. The  $M$ -sequence generators consisting of feedback shift registers are shown in Appendix 5.C [30]. The length of the  $M$ -sequence is  $2^N - 1$ , where  $N$  is the length of the shift register. In computer simulations, the data length of  $2^N$  is appropriate for the discrete Fourier transform (DFT). Therefore the test data is made by adding a “0” to the  $M$ -sequence: with this addition of a “0,” the number of “1s” and “0s” becomes equal. Importance of the balance will be mentioned later.



**FIGURE 5.35** Example of a digital transmission system for a computer simulation experiment.

In order for a digital computer to process signals, the signal must be sampled at discrete times. The sampling frequency must be high enough to satisfy the Nyquist sampling theorem (Section 2.4.1). From this viewpoint, simulation of a modulated signal at RF band is not appropriate. We then use the complex zero IF expression of the signal as well as the RF-circuit, for example, the band-pass filter. The complex zero-IF expression (Section 2.1.6) never loses generality if the bandpass signal is band-limited within a bandwidth smaller than the carrier frequency (a narrow band-pass signal). The nonlinearly modulated signal such as digital FM has, in principle, a spectrum that spreads into an infinite bandwidth, and hence the sampling theorem could not be strictly satisfied. However, the effect of the aliasing errors can be made sufficiently small by choosing an appropriately high sampling frequency, since the spectrum density decreases with increase in frequency.

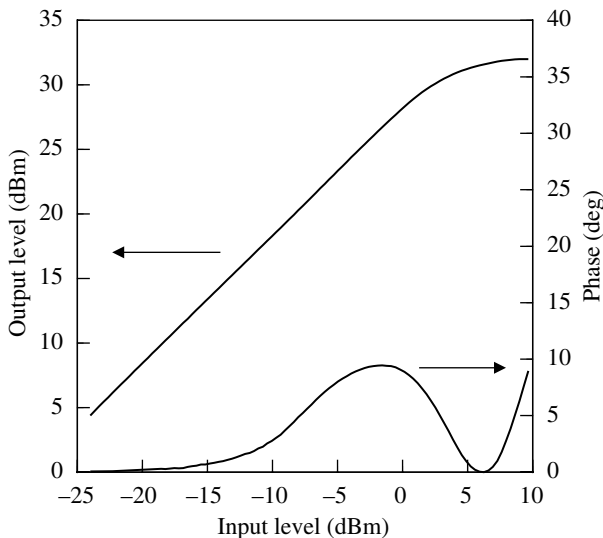
Filtering of a signal can be processed by convolution integral in time domain or multiplication in frequency domain. In many cases, frequency domain processing is efficient due to the fast DFT technique (Section 2.4.11). Attention must be paid to signal processing in the frequency domain. The DFT of a given signal block assumes a repetition of the signal block. The discontinuity of the signal from the end to the beginning of the signal block must be avoided: the effect of the discontinuity appears especially in unwanted spectrum spread and distortion of a signal at the beginning and the end of the signal block. In order to solve this problem, window function techniques are, in general, applied to the signal during processing. For a digitally modulated signal with linear modulation, the discontinuity between the signal at the end and the beginning of the signal block is not different from that between each data symbol and hence it is not a problem. For a digital FM signal, however, it becomes a problem. This problem can be simply avoided by balancing the plus and minus excursion of the modulating signal. This is the reason why balancing the “1” and “0” for a test data is previously recommended.

The effects of nonlinear distortion in an amplifier can be treated with the complex zero-IF signal, as far as the fundamental frequency component is concerned. As discussed in Appendix 5.A, the odd-order distortion for linear modulation is

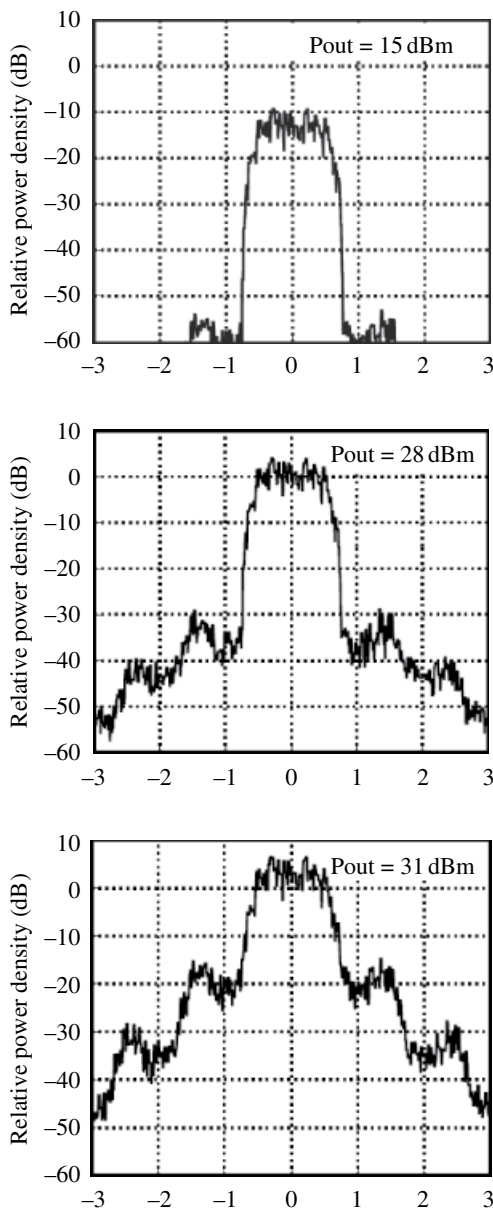
important. The output versus input relation of the amplifier, that is, the output signal envelope and phase shift as a function of the input signal level, called AM–AM conversion and AM–PM conversion, respectively, must be described. For example, Figure 5.36 shows a measured AM–AM and AM–PM conversion characteristics of a 900MHz band power amplifier module. To obtain the characteristics, we can use a network analyzer with variable output power. A computer-simulated power spectrum of a QPSK signal at the output of the power amplifier is shown in Figure 5.37. The first, second, and third out-of-band radiation components correspond to the third-, fifth- and seventh-order distortion, respectively. The average power efficiency can also be evaluated for a modulated signal, if the power efficiency is described as a function of the envelope of the input (unmodulated) signal.

The channel is simulated with a linear filter, which is either time-invariant or time-variant (fading). Co-channel or adjacent channel interference can be evaluated by adding the interfering signals. There are two ways to simulate the Rayleigh flat fading: a way is to add many signals with different amplitude and frequencies, which are distributed over the range of Doppler frequency. The principle of this method is described in Section 4.3. Prime number frequencies are desirable in order to generate a randomized output signal. The other way is to use quadrature modulation as shown in Figure 5.38. Principles of fading generator are discussed in Section 2.2.5. The power transfer function of the noise filter is given by the received power spectrum due to fast fading, for example, Equation 4.15. The frequency selective fading is simulated using a time-varying filter whose amplitude and/or delay characteristics depend on frequency.

The noise is simulated by a random number generator, which has a required stochastic characteristic, for example, additive white Gaussian noise. The characteristic

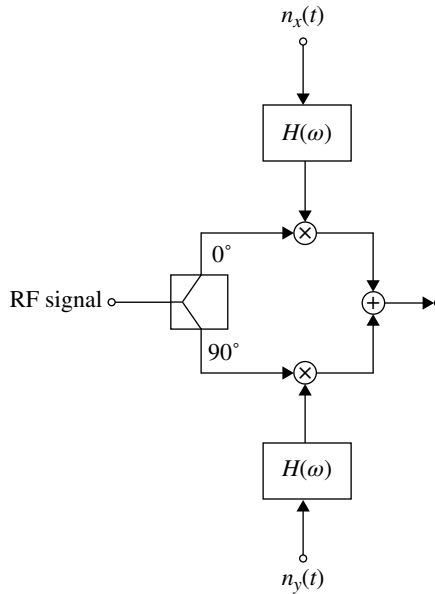


**FIGURE 5.36** AM–AM and AM–PM conversion characteristics of a 900MHz band power amplifier module (FMC-80802-20).



**FIGURE 5.37** Computer-simulated spectra of  $\pi/4$  shifted QPSK at the output of a quasi-linear power amplifier. Square-root of Nyquist-I filter with roll-off factor of 0.5 is used.

of the bandpass filter is described with an equivalent low-pass transfer function or impulse response. The operation of a demodulator can be described as discussed in Section 5.5. Sample and decision circuits including a clock recovery circuit can be described according to the actual system.



**FIGURE 5.38** Rayleigh fading generator.  $n_x(t)$  and  $n_y(t)$  are independent white Gaussian noise.

Many other parts of a digital transmission system, such as a diversity receiver, an automatic equalizer, and a carrier recovery circuit can be incorporated by mathematically describing their operation on the signal. The measuring instruments such as an oscilloscope for monitoring eye diagrams at various stages of the system, a spectrum analyzer, and an error counter are easily simulated. Implementing a subroutine library for each part of the transmission system is highly recommended, since we can easily set up our own system by picking up appropriate parts of the system. Importance of software simulators for digital communication systems will increase, since complexity of systems will increase and computing power is improving.

The topics discussed earlier fall in a category that is so called physical layer techniques. Performance evaluation of a total cellular system is also important. To do this performance evaluation, many base stations and terminals should be deployed and desired signal as well as undesired signal that interfere with the desired signal should be simulated taking into consideration radio wave propagation conditions. In this occasion, if the number of base stations and mobile stations is not sufficiently high, then the degree of interference in the cells at the edge of the service area is much lower than those in an inner cell. To avoid this kind of effect in performance evaluation, only inner cells are considered excluding the edge cells. However, the simulation time becomes long to get a sufficiently accurate result since stations of a smaller number are counted in the evaluation. To combat the problem, the wrap around system model is used, where the edge cells are connected together as if cells are deployed on a surface of a ball.



If we simulate a cellular system with many stations, taking all of the physical layer operations results in an intolerable running time or program size even with an advanced computer. To mitigate the difficulty in the simulation, physical level operation is replaced by a table look up method, where physical parameters (e.g., desired and undesired signal power and noise power) required for measurement of performances (e.g., throughput and packet error rates) are obtained in advance for the system level simulations. For a better accuracy of performance simulation, the generation of the table content becomes important. For example, in a physical layer technique, an adaptive array antenna system, which suppresses interference signal or get diversity effect, is included in a system level simulation. In this case, the required signal to interference and noise power ratio is theoretically given by Equation 7.115 on a basis of the minimum mean square error criterion.

## APPENDIX 5.A DISTORTION OF MODULATED SIGNAL APPLIED TO A NONLINEAR CIRCUIT

Let us express the input–output relation of the nonlinear circuit by series expansion as follows:

$$z(t) = a_1 y(t) + a_2 y^2(t) + a_3 y^3(t) \dots \quad (5.A.1)$$

where  $y(t)$  and  $z(t)$  are the input and output signal waveforms, respectively. The distortion is represented by the higher order components.

For a linear modulation, we denote

$$y(t) = A(t) \cos(\omega_c t)$$

and for a constant envelope modulation

$$y(t) = A_0 \cos\{\omega_c t + \phi(t)\}$$

We assume the modulated signal bandwidth is less than carrier frequency,  $\omega_c$ . From this assumption, the spectra of higher order distortion signals are assured not to overlap with each other. From Equation 5.A.1 and the fact that  $(\cos \omega_c t)^{2n}$  has no component of  $\cos \omega_c t$ , we can see that the even-order distortion has no frequency component at the frequency of the input signal  $y(t)$ . Thus, the even-order distorted signal can be eliminated by applying a band-pass filter, which is tuned to the carrier frequency.

For the constant envelope modulation signal, the carrier frequency component  $z_{\omega_c}(t)$  becomes

$$z_{\omega_c}(t) = A'_0 \cos\{\omega_c t + \phi(t)\}$$

where

$$A'_0 = a_1 A_0 + \frac{3}{4} a_3 A_0^3 + \frac{10}{16} a_5 A_0^5 + \dots$$

Thus, we get a distortion-free signal at the output of the bandpass filter.  
For the linear modulation signal

$$z_{\omega_c}(t) = A'(t) \cos(\omega_c t)$$

where

$$A'(t) = a_1 A(t) + \frac{3}{4} a_3 A^3(t) + \frac{10}{16} a_5 A^5(t) + \dots$$

The signal is distorted due to the odd-order components. The spectrum of the signal  $A^n(t)$  is given by  $n$ -fold convolution integral of the spectrum of  $A(t)$  (Eq. 2.40). Thus, its spectrum spreads  $n$  times wider than that of the input signal.

## APPENDIX 5.B DERIVATION OF THE EXPECTED GAUSSIAN NOISE POWER FOR FREQUENCY DISCRIMINATOR

Denoting  $a(t) = 1/A(t)$  and

$$b(t) = n_y(t) \cos \varphi(t) - n_x(t) \sin \varphi(t) \quad (5.B.1)$$

we have the demodulated noise from Equation 5.45 as

$$c(t) = a(t) b(t)$$

Then, we get the expected noise power at the output of the post-demodulation filter as

$$\begin{aligned} \langle N_g(t) \rangle &= \left\langle \left\{ \int_{-\infty}^{\infty} h_d(t-s) c(s) ds \right\}^2 \right\rangle \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_d(t-s_1) h_d(t-s_2) \langle c(s_1) c(s_2) \rangle ds_1 ds_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_d(t-s_1) h_d(t-s_2) a(s_1) a(s_2) \langle b(s_1) b(s_2) \rangle ds_1 ds_2 \end{aligned} \quad (5.B.2)$$

where  $h_d(t)$  is the total impulse response of the differentiating circuit and the post-demodulation filter. Using Equations 5.B.1, and 2.76, we get

$$\begin{aligned} &\langle b(s_1) b(s_2) \rangle \\ &= N_0 \int_{-\infty}^{\infty} g(s_1 - \tau) g(s_2 - \tau) \cdot \{ \cos \varphi(s_1) \cos \varphi(s_2) + \sin \varphi(s_1) \sin \varphi(s_2) \} d\tau \end{aligned} \quad (5.B.3)$$

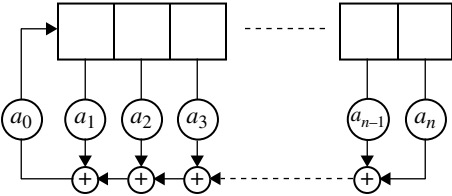
With Equations 5.B.2 and 5.B.3, we have

$$\begin{aligned} \langle N_g(t) \rangle &= N_0 \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} h_d(t-s) g(s-\tau) a(s) \cos \varphi(s) ds \right\}^2 \\ &\quad + \left\{ \int_{-\infty}^{\infty} h_d(t-s) g(s-\tau) a(s) \sin \varphi(s) ds \right\}^2 d\tau \\ &= N_0 \int_{-\infty}^{\infty} \left[ \left\{ h_d(t) * [g(t-\tau) a(t) \cos \varphi(t)] \right\}^2 \right. \\ &\quad \left. + \left\{ h_d(t) * [g(t-\tau) a(t) \sin \varphi(t)] \right\}^2 \right] d\tau \end{aligned}$$

APPENDIX 5.C M-SEQUENCE GENERATOR

M-sequence is generated with the feedback shift register [30]. The tap coefficients are given by  $(a_0, a_1, \dots, a_n)$  in the following table.

Number of stage, $n$	Period $(2^n - 1)$	$a_0, a_1, \dots, a_n$
2	3	111
3	7	1011
4	15	10011
5	31	100101
6	63	1000011
7	127	10001001
8	255	100011101
9	511	1000010001
10	1023	10000001001
11	2047	1000001010011
13	8191	10000000011011
14	16383	100010001000011
15	32767	100000000000011



## REFERENCES

- [1] B. P. Lathi, *Modern Digital and Analog Communication Systems*, Holt, Reinhart and Winston, New York, (1983).
- [2] J. G. Proakis, *Digital Communication*, 3rd ed McGraw-Hill, New York, (1995).
- [3] M. Schwartz, *Information Transmission, Modulation and Noise*, 3rd ed, McGraw-Hill, New York, (1980).
- [4] R. W. Lucky, J. Saltz, and E. J. Weldon, Jr., *Principles of Data Communication*, McGraw-Hill, New York, (1968).
- [5] W. R. Bennet, and J. R. Davey, *Data Transmission*, McGraw-Hill, New York, (1965)
- [6] M. Schwartz, W. R. Bennet, and S. Stein, *Communication Systems and Techniques*, McGraw-Hill, New York, (1966).
- [7] J. B. Anderson, T. Aulin, and C.-E. Sundberg, *Digital Phase Modulation*, Plenum Press, New York, (1986).
- [8] K. Fahrner, *Digital Communications, Satellite/Earth Engineering*, Prentice-Hall, Englewood Cliffs (NJ), (1983).
- [9] Baker PA. Phase-modulation data sets for serial transmission at 2000 and 2499 bits per second. In: AIEEE Trans, Part I (Communication and Electronics), July 1962. p 166–171.
- [10] F. G. Jenks, P. D. Morgan, and C. S. Warren, Use of four-level phase modulation for digital mobile radio. IEEE Trans Electromag Compat, EMC-14, 113–128 (1972).
- [11] K. Miyauchi, K. Izumi, S. Seki, and N. Ishida, Characteristics of an experimental guided millimeter-wave transmission system, IEEE Trans Commun, COM-20, 808–813 (1972).
- [12] T. Aulin, and C. E. Sundberg, An easy way to calculate power spectra for digital FM, IEE Proc F, 130, 519–525 (1983).
- [13] J. Namiki, Block demodulation for short radio packet. Trans IECE, 67-B, 54–61 (1984). [translated to English in Electronics and Communications in Japan, 67-B, 47–56, (1984).
- [14] R. F. Pawula, S. O. Rice, and J. H. Roberts, Distribution of the phase angle between two vectors perturbed by Gaussian noise, IEEE Trans Commun, COM-30, 1828–1841 (1982).
- [15] R. F. Pawula, Asymptotics and error rate bounds for M-ary DPSK, IEEE Trans Commun, COM-32, 93–94 (1984).
- [16] S. O. Rice, Noise in FM receivers, in *Time Series Analysis*, M. Rosenblatt, ed., John Wiley & Sons, Inc, New York, 1963.
- [17] J. E. Mazo and J. Saltz, Theory of error rates for digital FM, Bell Syst Tech J, 45, 1511–1535 (1966).
- [18] T. T. Tjhung and P. H. Wittke, Carrier transmission of binary data in a restricted band, IEEE Trans Commun Tech, COM-18, 295–304 (1970).
- [19] R. F. Pawula, On the theory of error rates for narrow-band digital FM, IEEE Trans Commun, COM-29, 1634–1643 (1981).
- [20] R. F. Pawula, Refinements to the theory of error rates for narrow-band digital FM, IEEE Trans Commun, COM-36, 509–513 (1988).
- [21] D. L. Schilling, E. Hoffman, and E. A. Nelson, Error rates for digital signals demodulated by an FM discriminator, IEEE Trans Commun Tech, COM-15, 507–517 (1967).

- [22] Y. Akaiwa and E. Okamoto, An analysis of error rates for Nyquist–and partial response–baseband-filtered digital FM with discriminator detection, *Trans IECE J66-B*, 534–541 (1983).
- [23] W. C. Jakes, ed., *Microwave Mobile Communications*, John Wiley & Sons, Inc, New York, (1974).
- [24] P. A. Bello and B. D. Nelin, The effect of frequency selective fading on the binary error probabilities of incoherent and differentially coherent matched filter receivers, *IEEE Trans Commun Syst*, CS-11, 170–186 (1963). [See also Corrections, *IEEE Trans. Communication Technology*, COM-12, 230 (December 1964).]
- [25] C. C. Bailey and J. C. Lindenlaub, Further results concerning the effect of frequency-selective fading on differentially coherent matched filter receivers, *IEEE Trans Commun Tech*, COM-16, 749–751 (1968).
- [26] K. Hirade, M. Ishizuka, and F. Adachi, Error-rate performance of digital FM with discriminator-detection in the presence of co-channel interference under fast Rayleigh fading environment, *Trans IECE*, E61, 704–709 (1978).
- [27] K. Hirade, M. Ishizuka, F. Adachi, and K. Ohtani, Error-rate performance of digital FM with differential detection in land mobile radio channels, *IEEE Trans Veh Tech*, VT-28, 204–212 (1979).
- [28] F. Adachi and J. D. Parsons, Error rate performance of digital FM mobile radio with postdetection diversity, *IEEE Trans Commun*, 37, pp. 200–210, (1989).
- [29] F. Adachi and K. Ohono, BER performance of QDPSK with postdetection diversity reception in mobile radio channels, *IEEE Trans Veh Tech*, 40, 237–249, (1991).
- [30] Kaneko H. *PCM Communication Technology*, Sanpou, (1976), (in Japanese).

This page intentionally left blank

---

# 6

---

## **DIGITAL MODULATION/ DEMODULATION FOR MOBILE RADIO COMMUNICATION**

Digital modulation has been used for a long time in analog FM mobile radio communication systems. The role of digital transmission in the analog FM mobile radio system is to send low-speed data signals to control the system, for example, paging a called mobile terminal and setting up and releasing of a dedicated channel. Spectrum efficiency of the digital modulation was not a key issue for such systems, since the data transmission rate is low: from several hundreds to thousands of bits per second.

The history of research and development of spectrum-efficient digital modulation techniques for mobile radio communication is not so long. It started in 1978 with the invention of spectrum-efficient digital FM by de Jager and Dekker. After their invention, many digital modulation methods were proposed in a very short period, described later in detail. The motivation for efficient digital modulation was digital voice communications.

This chapter describes some digital modulation methods in order of appearance: digital modulation for analog FM systems, spectrum-efficient digital FM, linear modulation, spread spectrum system, multicarrier transmission, and single-carrier frequency-division modulation. Research and development in digital modulation techniques for mobile radio communications are still in progress at the time of the writing of this book.

Special requirements for digital mobile radio communication are stable performance of bit error rate in fast-fading channels and cochannel interfering environments, low power consumption, as well as size and cost-effectiveness of equipment. Due to these special requirements, digital modulation/demodulation for mobile radio communication differs from that for other communication systems.

## 6.1 DIGITAL MODULATION FOR ANALOG FM MOBILE RADIO SYSTEMS

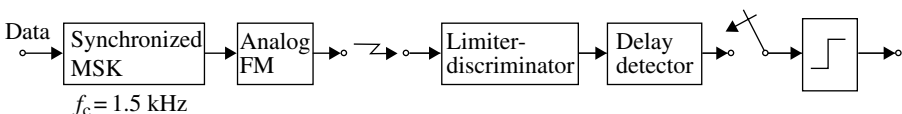
An important issue is how to incorporate the digital modulation/demodulation into an analog FM system. The easiest and actually adopted method is to transmit the data signal in the voice channel of the analog FM system. This is similar to data communication through wire-line analog public telephone channels. Since the spectrum of the voice signal covers frequencies from about 0.3 to 3 kHz, the voice channel for analog FM is designed not to pass a dc signal: a dc circuit is rather difficult to implement.

For transmission of a digital signal over the dc-blocked channel, we have two methods: one is to use a line code that has no dc component; the other is to use a modulated signal whose carrier frequency is in the middle of the voice band channel. The Manchester code described in Section 3.2.4 is useful for the former method. For the latter method, called the subcarrier system, minimum shift keying (MSK) with a carrier frequency around 1.5 kHz is sometimes used. The subcarrier system is shown in Figure 6.1. The carrier signal is synchronized to the data clock signal for convenience in demodulation of the MSK signal, whose carrier and data frequencies are comparable.

Another subcarrier system uses a data modem (modulator/demodulator), which was developed for wire-line public telephone channels [1]. The modem uses different modulation for different data transmission rates from 300 bps to 28.8 kbps. A high-speed modem employs sophisticated techniques including automatic channel equalization. For application of the data modem to the analog voice FM mobile radio channels, the highest data rate is limited to around 4800 kbps, due to the fading phenomenon in mobile radio channels.

## 6.2 CONSTANT ENVELOPE MODULATION

This type of modulation is advantageous for application to mobile radio communications because a saturated power amplifier, which is principally power efficient, can be employed as discussed in Section 5.2. Furthermore, a band-pass limiter circuit can be used at a receiver. The limiter circuit can easily be implemented to cope with fast-fading and wide dynamic range effects in mobile radio channels. Therefore, it is natural that many researchers have focused on constant envelope digital modulation. From the viewpoint of compact spectrum characteristics, a continuous-phase frequency shift keying (CPFSK) with a low modulation index becomes important in mobile radio channels.



**FIGURE 6.1** Subcarrier MSK signal transmission for an analog FM system.



### 6.2.1 MSK

Binary FM with rectangular (NRZ) pulse shaping and limiter–discriminator detection is the most primitive system. Rectangular pulse-shaped binary CPFSK with a modulation index of 0.5 is known as MSK or fast frequency shift keying (FFSK) [2, 3]. An MSK signal can be expressed as

$$s(t) = A_0 \cos \left( \omega_c t + a_n \frac{\pi}{2T} t' + \frac{\pi}{2} \sum_{j=1}^{n-1} a_j \right) \quad ((n-1)T \leq t \leq nT) \quad (6.1)$$

where  $t' = t - (n-1)T$ ,  $A_0$  is the amplitude,  $\omega_c$  is the carrier frequency,  $a_n$  takes the values  $\pm 1$  corresponding to the input digital signal, and  $T$  is the bit duration. The frequency becomes  $f_M = f_c - 1/4T$  (mark) and  $f_S = f_c + 1/4T$  (space), corresponding to the input data, where  $f_c = \omega_c / 2\pi$ .

The name MSK comes from the fact that the modulation index is minimum for a pair of FSK signals that are orthogonal, corresponding to the two-level signal. For orthogonal signals, we can apply coherent detection without interference between the two signals. The coherent detector for the MSK signal is shown in Figure 6.2. The low-pass (zonal) filter is intended to suppress the harmonic signal components. The noncoherent detector can be obtained if we place an envelope detector between the low-pass filter and the integrator.

At the end of a symbol time, the signal takes a phase from two groups in turn: one group includes  $+\pi/2$  and  $-\pi/2$ , and the other includes 0 and  $\pi$ , as shown in Figure 6.3. The phase constellation is the same as  $\pi/2$ -shifted BPSK. The signal phase moves on the circle with a constant speed.

Equation 6.1 can be rewritten as

$$s(t) = x_n(t) \cos \omega_c t - y_n(t) \sin \omega_c t \quad ((n-1)T \leq t \leq nT) \quad (6.2)$$

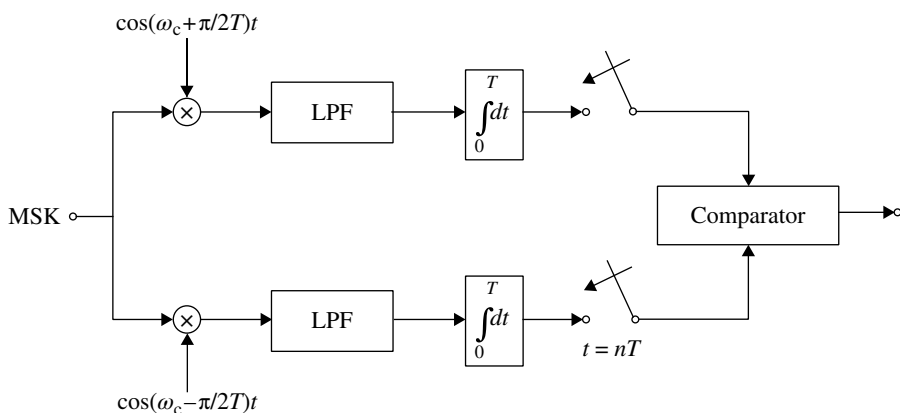
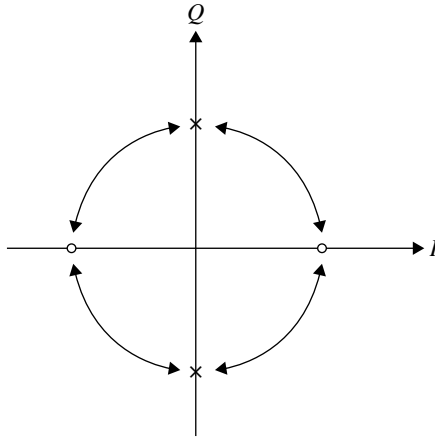


FIGURE 6.2 A coherent detector for MSK signals.



**FIGURE 6.3** Phase constellation of MSK signals.

where

$$x_n(t) = A_0 \cos\left(a_n \frac{\pi}{2T} t' + \frac{\pi}{2} b_{n-1}\right) \quad ((n-1)T \leq t \leq nT) \quad (6.3)$$

$$y_n(t) = A_0 \sin\left(a_n \frac{\pi}{2T} t' + \frac{\pi}{2} b_{n-1}\right) \quad ((n-1)T \leq t \leq nT) \quad (6.4)$$

and

$$b_{n-1} = \sum_{i=1}^{n-1} a_i \quad (6.5)$$

Using Equations 6.3–6.5 and expanding the time region to  $2T$ , we define

$$\begin{aligned} x'_{2m}(t) &\equiv x_{2m}(t) + x_{2m+1}(t) \\ &= -a_{2m} \sin\left(\frac{\pi}{2} b_{2m-1}\right) \sin\left(\frac{2\pi}{2T} t'_1\right) \quad (2mT \leq t \leq 2(m+1)T) \end{aligned} \quad (6.6)$$

where  $t'_1 = t - 2mT$  and define

$$\begin{aligned} y'_{2m}(i) &= y_{2m-1}(i) + y_{2m}(i) \\ &= a_{2m-1} \cos\left(\frac{\pi}{2} b_{2m-2}\right) \sin\left(\frac{\pi}{2T} t'_2\right) \quad ((2m-1)T \leq t \leq (2m+1)T) \end{aligned} \quad (6.7)$$

where  $t'_2 = t - (2m-1)T$

If we correspond 1 and 0 with 1 and  $-1$ , respectively, for  $a_n$ , we have the following expressions:

$$c_{2m} = c_{2m-1} \oplus \bar{d}_{2m} \quad (6.8)$$

$$c_{2m-1} = c_{2m-2} \oplus d_{2m-1} \quad (6.9)$$

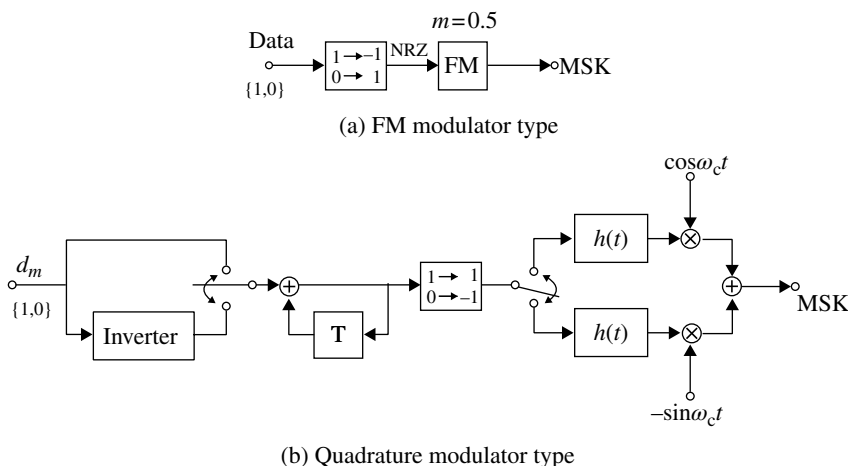
where  $\oplus$  means mod 2 addition,  $d_n = \{1, 0\}$  denotes the input data, and  $c_n$  is the differentially encoded data, which is inverted every even time slot. Thus, as well as a conventional FM modulator, we have an MSK signal modulator with quadrature modulation as shown in Figure 6.4. The impulse response is given by  $h(t) = \sin[(\pi/2T)t]$  for  $0 \leq t \leq 2T$  and  $h(t) = 0$  otherwise.

The power spectral density of MSK signals for random input data can be calculated by following the discussion for constant envelope modulation in Section 5.4.2. However, the easiest way is to follow the method for linear modulation discussed in Section 5.4.1, noting that the MSK signal can be generated with quadrature modulation. Assuming that the input data and therefore the inphase and the quadrature component signals are random, we have the power spectral density as

$$S(\omega) = \frac{(4/\pi)^2 T}{[1 - (2\omega T/\pi)^2]^2} \cos^2(\omega T) \quad (-\infty \leq \omega \leq \infty) \quad (6.10)$$

This is shown in Figure 6.5.

Before we go further, let us compare the inphase and quadrature waveforms of MSK, offset QPSK, and  $\pi/2$ -shifted BPSK in Figure 6.6. An MSK signal can be



**FIGURE 6.4** Modulators for MSK signals.

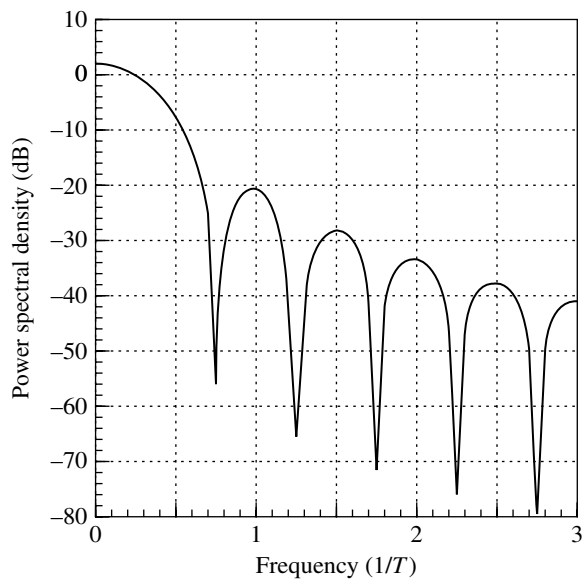


FIGURE 6.5 Power spectrum of MSK.

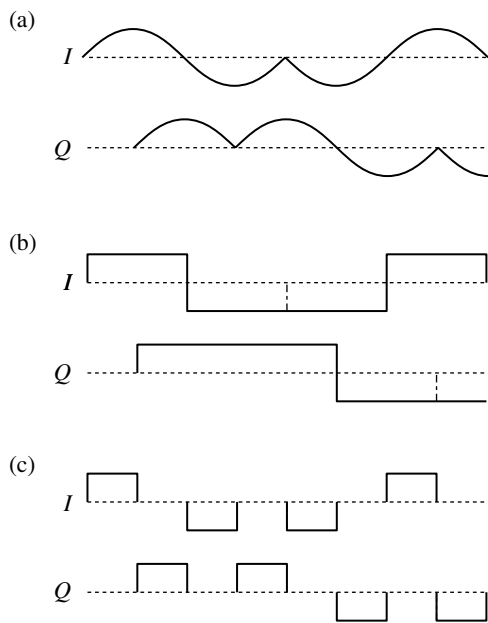
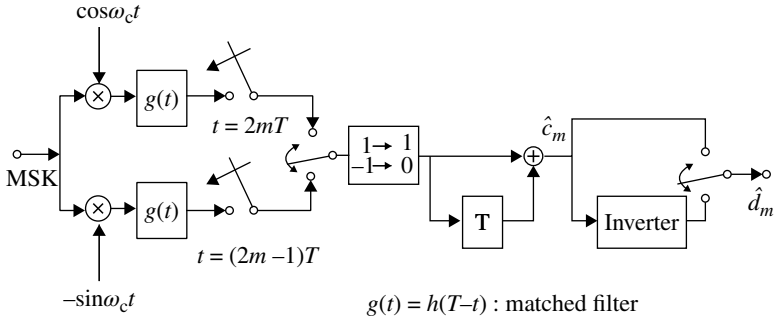


FIGURE 6.6 Comparison of the inphase and quadrature waveforms between (a) MSK, (b) offset QPSK, and (c)  $\pi/2$ -shifted BPSK.



**FIGURE 6.7** Coherent detector for MSK signal.

understood as a special case of offset QPSK, whose pulse waveform is  $h(t) = \sin(\pi t/2T)$  for  $0 \leq t \leq 2T$  and  $h(t) = 0$  for  $t < 0$  or  $t > 2T$ .

Expression of an MSK signal by Equations 6.2–6.9 and its modulator configuration in Figure 6.4b suggest a coherent detector as shown in Figure 6.7. This detector is different from the previous one (Fig. 6.2): the observation time is expanded to  $2T$  instead of  $T$ . The eye diagram of quadrature demodulated MSK is shown in Figure 6.8. A recovered carrier-phase slip by  $\pm 180^\circ$  never affects the received data because of the differential encoding embedded in the MSK modulation and the differential decoding at the receiver.

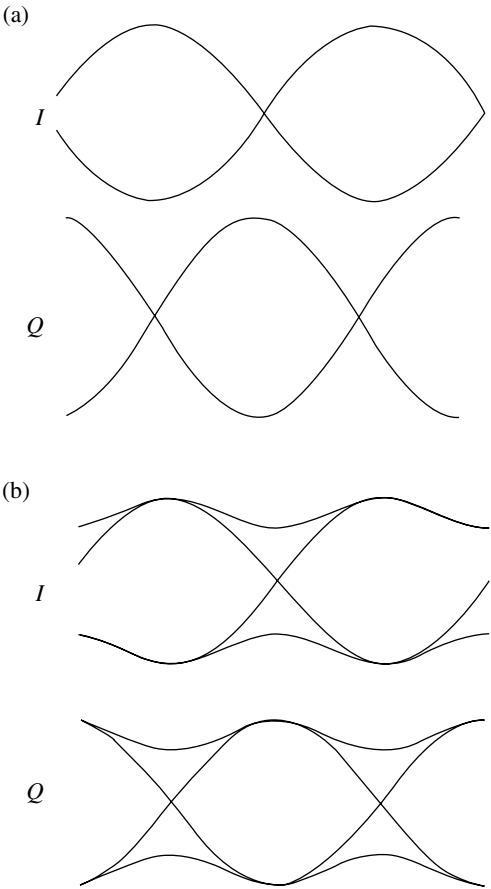
A carrier recovery circuit for an MSK signal is shown in Figure 6.9. By applying an MSK signal to a frequency doubler, we have discontinuous frequency components at  $2f_M$  and  $2f_S$ , where  $f_M = f_c - 1/4T$  and  $f_S = f_c + 1/4T$ . The existence of the discontinuous frequency components in the frequency-doubled signal can be supported by confirming that  $|C_\alpha| = 1$  with  $M = 2$ ,  $L = 1$ , and  $h(\equiv m) = 1$  (see Section 5.4.2), where the modulation index is doubled from frequency doubling. Each component is entered into a resonator to reduce noise components. The principle of operation of the carrier recovery is in the frequency quadruplication. A phase ambiguity of  $\pm\pi/2$  or  $\pm\pi$  occurs in the frequency quadri-divider. In order to cope with the phase ambiguity of  $+\pi/2$  and  $-\pi/2$ , differential encoding is introduced. The transmitter and receiver are shown in Figure 6.10. The encoded signal  $c_n$  and the input data  $d_n$  are expressed as

$$d_n = c_{n-1} \oplus c_n \quad \text{or equivalently} \quad c_n = c_{n-1} \oplus d_n$$

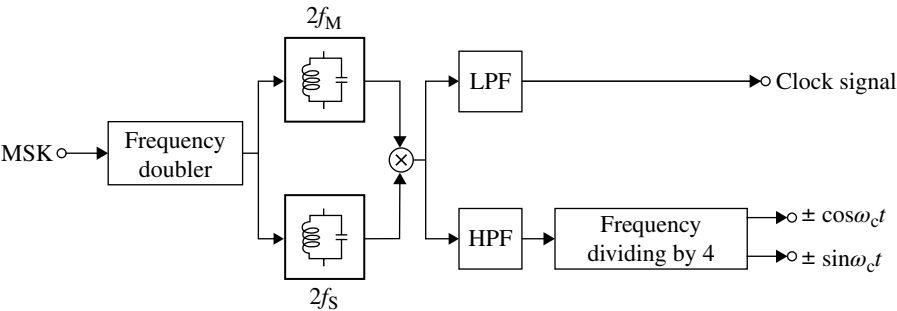
From the characteristics of the MSK signal, the phase shift during two symbol times ( $2T$ ) becomes  $180^\circ$  and  $0^\circ$ , corresponding to  $c_{n-1} \oplus c_n$  equaling 0 and 1, respectively. Here  $c_{n-1} \oplus c_n = 0$  means  $c_{n-1} = c_n$  (successive 1s or 0s) and  $c_{n-1} \oplus c_n = 1$  means  $c_{n-1} \neq c_n$ . In an error-free transmission, we have

$$\overline{d_{2n}} = \overline{c_{2n-1}} \oplus \overline{c_{2n}} = \overline{I_{2n-2}} \oplus \overline{I_{2n}} = \hat{d}_{2n} \quad (6.11a)$$

$$\overline{d_{2n+1}} = \overline{c_{2n}} \oplus \overline{c_{2n+1}} = \overline{Q_{2n-1}} \oplus \overline{Q_{2n+1}} = \hat{d}_{2n+1} \quad (6.11b)$$



**FIGURE 6.8** Eyediagram of a matched filtered MSK signal with coherent detection: (a) non-filtered and (b) matched filtered.

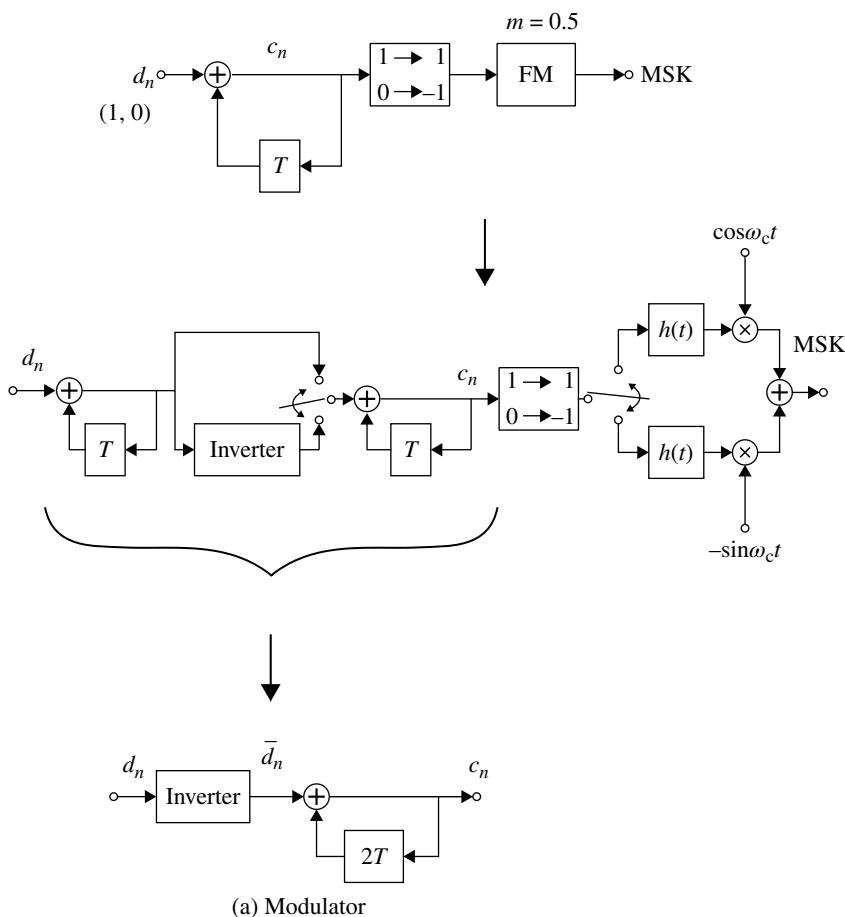


**FIGURE 6.9** Carrier recovery circuit for MSK signals.

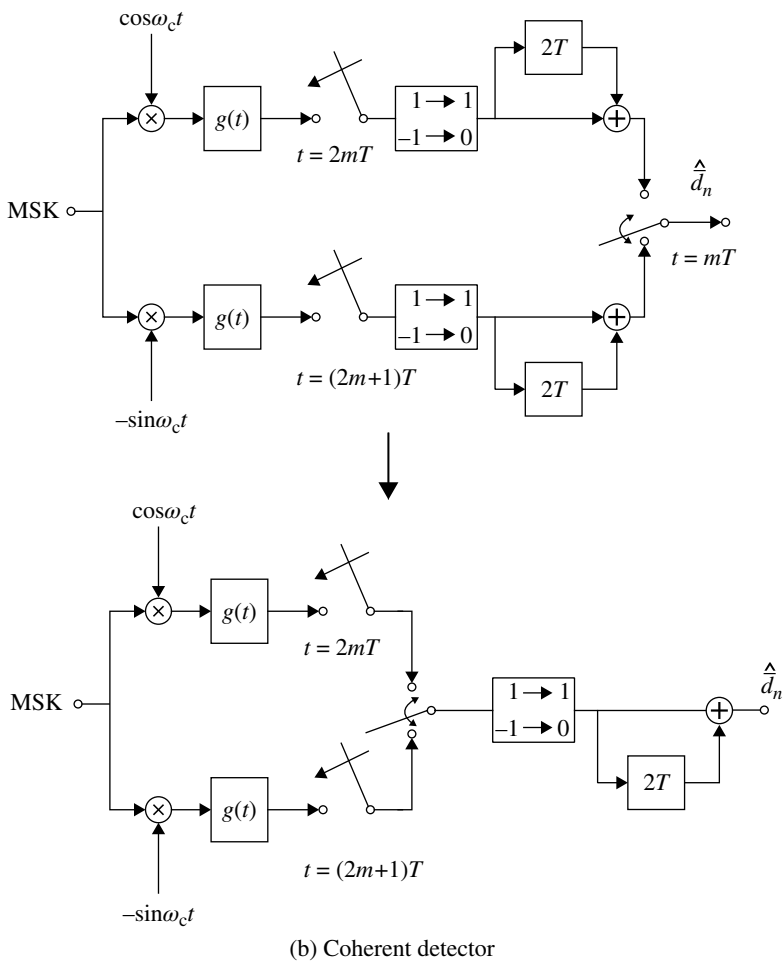
where  $I_n$  and  $Q_n$  are the decided inphase and quadrature data signals decided at the receiver and  $\hat{d}_n$  are the received data signals. We can see that the phase ambiguity of  $\pm\pi/2$ , as well as  $\pm\pi$ , has no effect on the received data, since inversion of  $I_n$  and  $Q_n$  ( $180^\circ$  ambiguity) and/or interchange of  $I_n$  and  $Q_n$  ( $\pm 90^\circ$  ambiguity) never changes  $\hat{d}_n$  in Equations 6.11a and 6.11b.

A coherent detector that requires no differential encoding was proposed in Ref. [4] (Fig. 6.11). The symbol  $\oplus$  denotes mod 2 addition. The principle operation of the coherent detector is the same as that shown in Figure 6.7. The phase ambiguity of  $\pi/2$  and  $-\pi/2$  is solved by the clock signal-assisted carrier recovery system. The effect of carrier signal phase slip is compensated by the recovered clock signal.

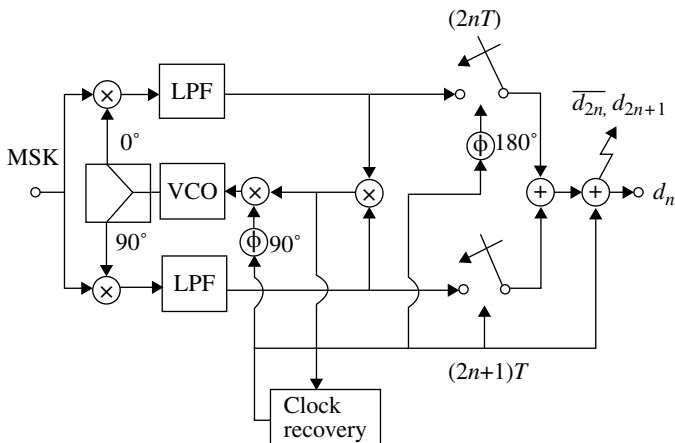
Another coherent detector [5] for MSK signals is shown in Figure 6.12. The circuit consists of only one path, similarly to a coherent detector for BPSK signals. The recovered local signal, which is synchronized to the mark or space frequency, is



**FIGURE 6.10** Modulator and coherent detector for MSK signals with differential encoding.

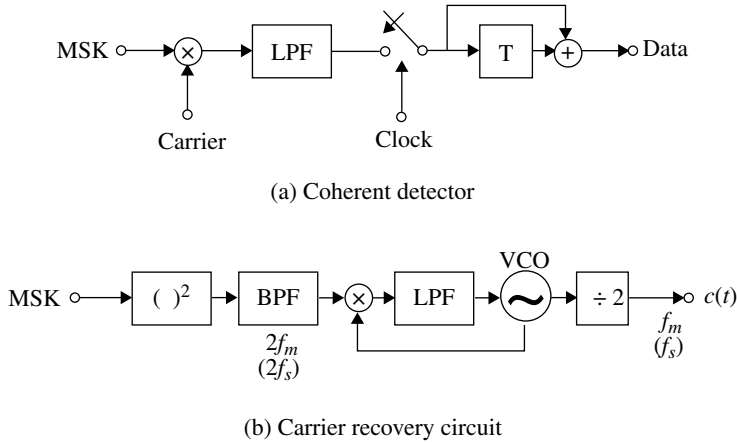


**FIGURE 6.10** (Continued)



**FIGURE 6.11** Coherent detector for MSK signal, which requires no differential encoding [4]. Reproduced by permission of Copyright © 1981, IEICE.





**FIGURE 6.12** Coherent detector for MSK signals with deviated-frequency-locking scheme [5].

used for demodulation (deviated-frequency method). The relative phase of the MSK signal to the local signal phase is  $0^\circ$  or  $180^\circ$ .

The other method for carrier recovery of the MSK signal is described in [2], where the signal required for the correlation receiver is obtained together with the carrier signal.

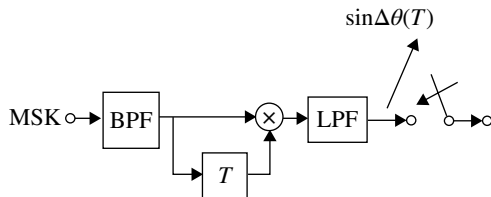
The error rate of an MSK signal with coherent detection is different depending on the type of coherent detector. A coherent detector that observes a received signal for two symbol periods (Fig. 6.7) shows a superior performance by 3 dB in signal to noise power ratio compared to a coherent detector that observes for one symbol period (Fig. 6.2). This is because the former uses an orthogonal and the latter an antipodal signaling system. The optimum error rate performance is obtained by matched filtering and coherent detection with 2-symbol-time observations, as discussed in Section 3.3. Eye diagrams of non-filtered and matched filtered MSK signals with coherent detection are shown in Figure 6.8.

Two kinds of differential detection of MSK signals are known (Fig. 6.13). The phase shift during one and two symbol times is  $(\pi/2)a_n$  and  $(\pi/2)(a_n + a_{n+1})$  ( $a_n = \pm 1$ ), respectively. By adjusting the carrier-phase difference between the two paths in the detector, the signal at the output of the zonal low-pass filter becomes, for a one-bit delay detector,

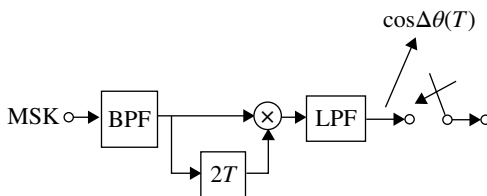
$$s_d(nT) = \sin\left(\frac{\pi}{2}a_n\right) = \begin{cases} 1 & (a_n = 1) \\ -1 & (a_n = -1) \end{cases} \quad (6.12a)$$

and, for a two-bit delay detector,

$$s_d(nT) = \cos\left[\frac{\pi}{2}(a_n + a_{n+1})\right] = \begin{cases} 1 & (a_n + a_{n+1} = 0) \\ -1 & (a_n + a_{n+1} = \pm 2) \end{cases} \quad (6.12b)$$



(a) One-bit delay detector



(b) Two-bit delay detector

**FIGURE 6.13** Differential detector for MSK signals.

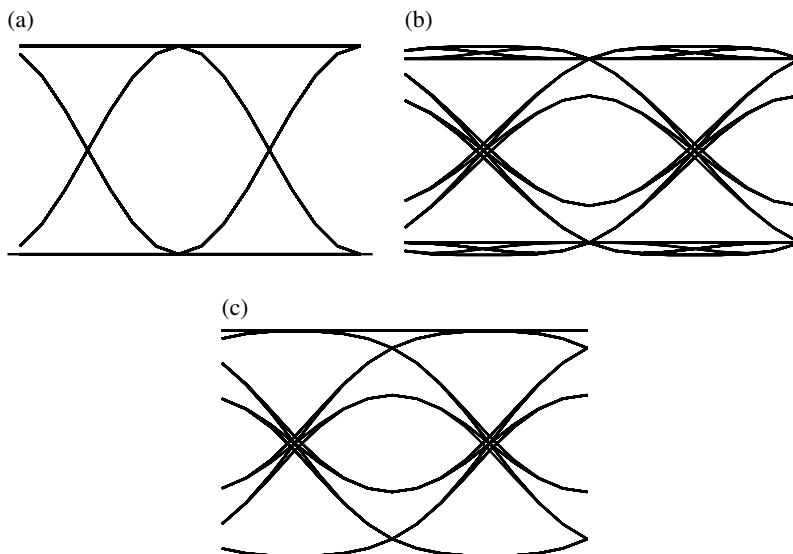
Differential encoding is necessary for the two-bit delay detector, since the differential detection is equivalent to differential decoding on the data signal.

Contrary to coherent detection, differential detection of the matched filtered MSK signal shows intersymbol interference (Fig. 6.14). The matched filtered signal  $s(t)$  becomes  $s(t) = \pm h(t) * h(2T - t)$ , where  $h(t) = \sin(\pi t/2T)$  for  $0 \leq t \leq 2T$  and  $h(t) = 0$  otherwise.  $s(2T)$ , the inphase signal, becomes  $T$  at the sampled time. At this instant,  $s(T)$ , the quadrature component, takes a value of  $0$ ,  $+2T/\pi$ , and  $-2T/\pi$ . Thus, the phases of the matched filtered MSK signals scatter by  $0$  or  $\pm \tan^{-1}(2/\pi)$ .

Optimization of the receiver filter for differential detection of MSK is discussed in [6]. A differential detection system for MSK with nonredundant error correction capability is proposed in [7].

The other method for demodulation of MSK signals is limiter–discriminator detection. An integrate-and-dump (I&D) filter is exclusively used as a postdetection filter, since it shows no intersymbol interference when receiving an NRZ signal, which is obtained at the demodulator output. The I&D filter is not the matched filter for postdetection, since the noise spectrum is not flat at the output of the limiter–discriminator.

The matched filter receiver for either coherent or differential detection of MSK signals is not actually recommended for application to mobile radio communications. The reason is that the matched filter offers poor channel selectivity performance, which is imperative to compensate for the near-far problem (Section 4.5) in mobile radio communications.

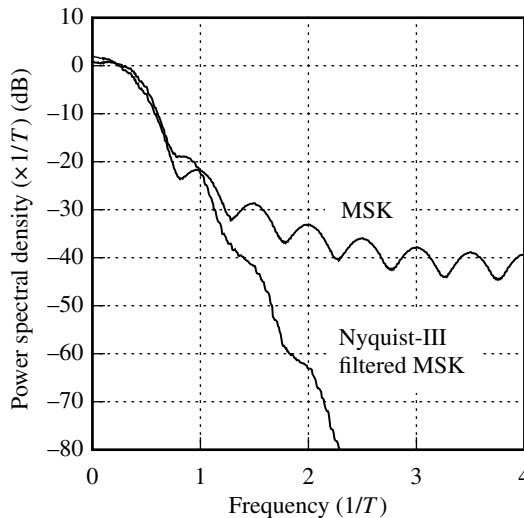


**FIGURE 6.14** Eye diagram of an MSK with differential detection: (a) nonfiltered, (b) matched filtered, and (c) matched filtered and amplitude limited.

Better channel selectivity is obtained by using a band-pass filter, which shows a steeper transfer function, such as a Gaussian filter. Assuming a general band-pass filter, the error rates for MSK with differential detection, as well as limiter-discriminator detection are analyzed in [8], where calculated results are presented for a Gaussian band-pass filter.

Even when a band-pass filter with good channel selectivity is adopted, MSK is not to be recommended for mobile radio communication, since its spectrum contains high out-of-band radiation: channel separation must be wide to get sufficient channel selectivity. In order to control the out-of-band radiation of MSK, pulse shapes have been proposed for the premodulation signal [9–11]. However, the results are not adequate for application to mobile radio communications, since the band-limited pulse waveforms are not stringent. The effect of strict band-limited pulse waveforms on MSK-type signals is demonstrated in Figure 6.15. The out-of-band radiation is decreased by using the Nyquist III roll-off filter (Section 3.1.3), where the base-band signal spectrum is rigorously band limited within  $|\omega| \leq (1 + \alpha) \pi/T$ , where  $T$  is the bit duration and  $\alpha$  is the roll-off factor. The characteristic of MSK, in which the modulated signal takes fixed phases from  $0$ ,  $\pm\pi/2$ , and  $\pi$ , is preserved by using the Nyquist III filter.

We have discussed MSK in detail, since it presents a good reference system for other constant envelope digital modulations.



**FIGURE 6.15** Effect of a band limitation of premodulation signal on spectrum for an MSK-type signal.

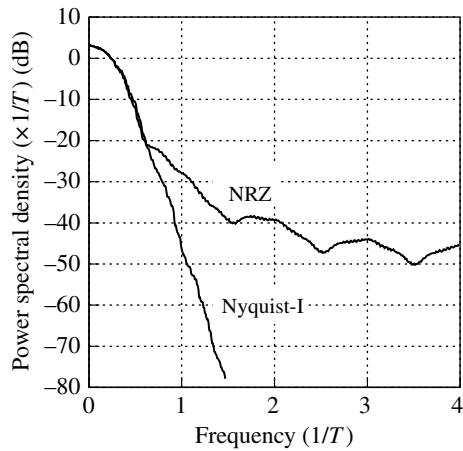
## 6.2.2 Partial-Response Digital FM

Partial-response digital FM was proposed many years ago, for example, duobinary FM by Lender [12]. However, rigorous band limitation on the premodulation signal was not adopted. In 1978, de Jager and Dekker proposed a pioneering digital FM system called tamed frequency modulation (TFM) [13], which meets the requirements for application to mobile radio communications: the out-of-band radiation is drastically decreased. Their success is attributed to a partial-response digital FM system, in which band limitation of the premodulation baseband signal is strict and the modulation index is low. The proposal of TFM stimulated many researchers to develop other spectrum-efficient digital FM systems.

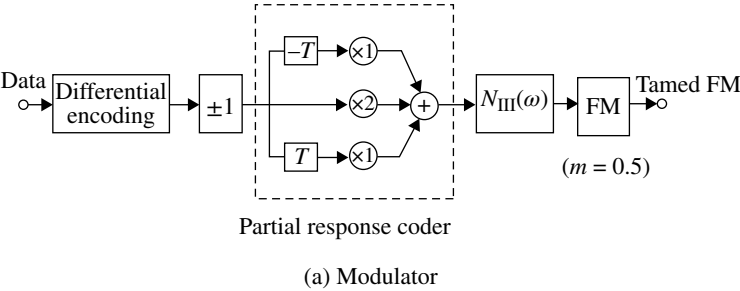
**6.2.2.1 Duobinary FM** In this system, duobinary coding is employed for FM with limiter–discriminator detection. The baseband signal takes three levels due to the duobinary coding (Section 3.2.6). Figure 6.16 shows the spectra of duobinary FM for NRZ and Nyquist I pulse shaping. The out-of-band radiation is decreased due to the band limitation of the pre-modulation signal by the Nyquist I filter. The error rate performance of noncoherent demodulation of duobinary coded MSK and TFM in fading mobile radio channels is described in [14].

**6.2.2.2 TFM** The block diagram of TFM is shown in Figure 6.17. A class II partial-response (Section 3.2.6) system is introduced to digital FM with modulation index of 0.5, and the Nyquist III pulse shaping (Section 3.1.3) is employed. The transfer function becomes

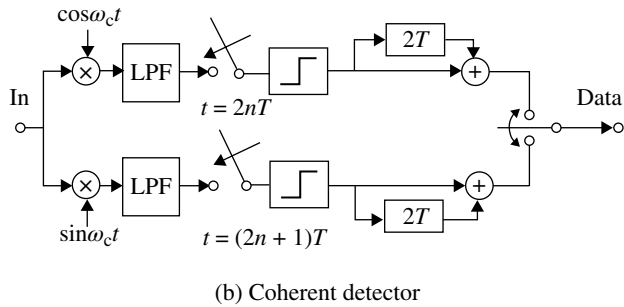
$$H(\omega) = \cos^2\left(\frac{\omega T}{2}\right) N_{\text{III}}(\omega) \quad (6.13)$$



**FIGURE 6.16** Power spectral densities of duobinary FM for NRZ and Nyquist I pulse shaping (roll-off factor of zero).



(a) Modulator



**FIGURE 6.17** Block diagram of a tamed FM system.

where  $T$  is the bit duration and  $N_{\text{III}}(\omega)$  denotes the Nyquist III transfer function.  $N_{\text{III}}(\omega)$  is given as

$$N_{\text{III}}(\omega) = \frac{\omega T / 2}{\sin(\omega T / 2)} N_{\text{I}}(\omega)$$

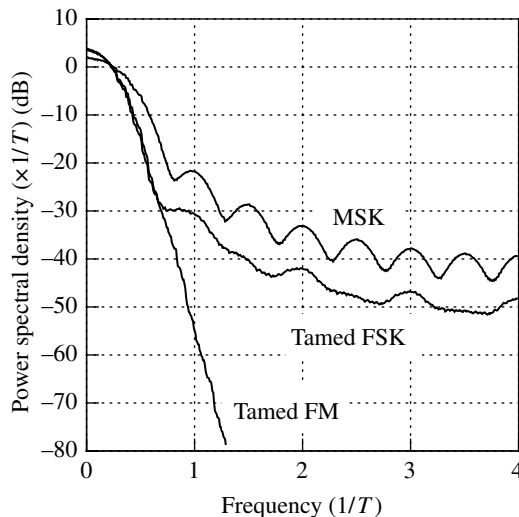
where  $N_{\text{I}}(\omega)$  denotes the Nyquist I transfer function. The term  $\cos^2(\omega T / 2)$  represents class II partial-response signaling and is given by the Fourier transform of its impulse response, that is,  $\delta(t+T) + 2\delta(t) + \delta(t-T)$ .

The phase shift during one-bit duration is given as

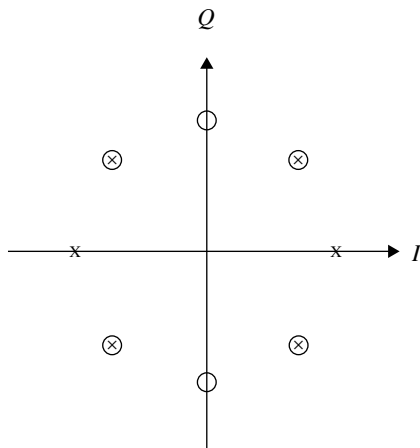
$$\begin{aligned} \Delta\theta(nT) &= \theta(nT+T) - \theta(nT) \\ &= \frac{\pi}{2} \left( \frac{1}{4}a_{n-1} + \frac{1}{2}a_n + \frac{1}{4}a_{n+1} + 1 \right) \end{aligned} \quad (6.14)$$

where the data symbols  $a_n$  take values +1 or -1.

The power spectral density of TFM together with MSK is shown in Figure 6.18. Tamed FSK is given by replacing the Nyquist III filter with an NRZ pulse-shaping filter. When we apply the TFM to the conventional FM channels, where the channel separation is 25 kHz and the adjacent channel interference is around -60 dB, we can accommodate digital transmission data rates of 16 kbps. At the time when TFM was invented, a one-chip voice coder/decoder (CODEC) with adaptive delta modulation (ADM; Section 7.7.2) was available. The combination of 16 kbps ADM voice coding and TFM was a historical landmark for digital voice communication through mobile radio channels.



**FIGURE 6.18** Power spectral densities of tamed FM and MSK signals.



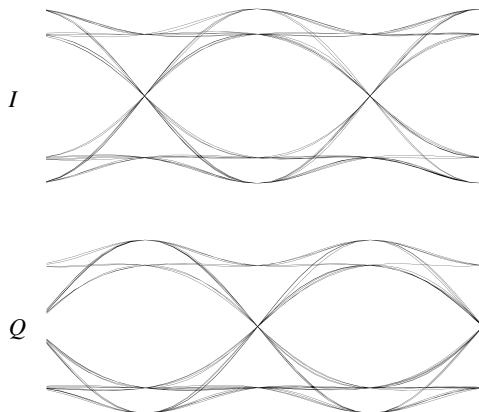
**FIGURE 6.19** Phase constellation of tamed FM.

The TFM takes one of fixed phases, namely,  $(\pi/8)(a_{n-1} + 2a_n + a_{n+1})$  ( $a_n = \pm 1$ ), at each symbol time (Fig. 6.19) since the modulation index is chosen to be 0.5 and Nyquist III pulse shaping is employed. The signal phase is taken in turn from the two signal groups marked by the cross and circle in Figure 6.19.

By comparing the phase constellations for TFM and for MSK (Fig. 6.3), we can see that the TFM signal scatters due to the introduction of partial-response signaling. Even with the scattering of the signal phases, an eye opening is obtained for the inphase and quadrature components of TFM, as shown in Figure 6.20. Therefore, the coherent demodulation system, which is similar to that of MSK (Fig. 6.10), can be applied to TFM. The difference between the two systems resides in the low-pass filter. In contrast to MSK, the waveform of inphase or quadrature signal component of TFM is not time limited within two symbol times. Also, it never meets Nyquist's first criterion. Thus, the receiver with symbol-by-symbol decision does not yield optimum error rate performance.

The optimum receiver filter for quadrature coherent demodulation of digital FM with modulation index of 0.5 is investigated in [15]. Their results for TFM show that the performance degradation in terms of  $E_b/N_0$ , compared to MSK, is 0.75 and 1.2 dB at error rates of  $10^{-3}$  and  $10^{-6}$ , respectively. From the viewpoint of actual applications, the receiver filter must be designed taking into consideration the channel selectivity as well as the error rate performance.

Differential detection of TFM shows poor error rate performance [14] due to intersymbol interference in the demodulated signal. Limiter-discriminator detection of TFM results in a demodulated signal with five levels due to the class II partial-response coding. Transmitted data can be recovered by decoding the five levels into two levels (see Section 3.2.6). In this case, the precoding  $H_p(z) = 1/(1 + 2z^{-1} + z^{-2}) \pmod{2}$  is assumed. Intersymbol interference, which is in excess from partial-response coding, appears because Nyquist III filtering is used instead of Nyquist I filtering.



**FIGURE 6.20** Eye diagram of tamed FM with coherent detection.

The experimental and computer-simulated error rate performance of TFM is described in [16–18].

**6.2.2.3 Generalized TFM** In this system [19], generalized partial-response (Section 3.2.6) signaling is used for the premodulation signal. The phase shift during a bit duration is given as

$$\begin{aligned}\Delta\theta(nT) &= \theta(nT+T) - \theta(nT) \\ &= \frac{\pi}{2}(Aa_{n-1} + Ba_n + Aa_{n+1}) \quad (a_n = \pm 1)\end{aligned}\tag{6.15}$$

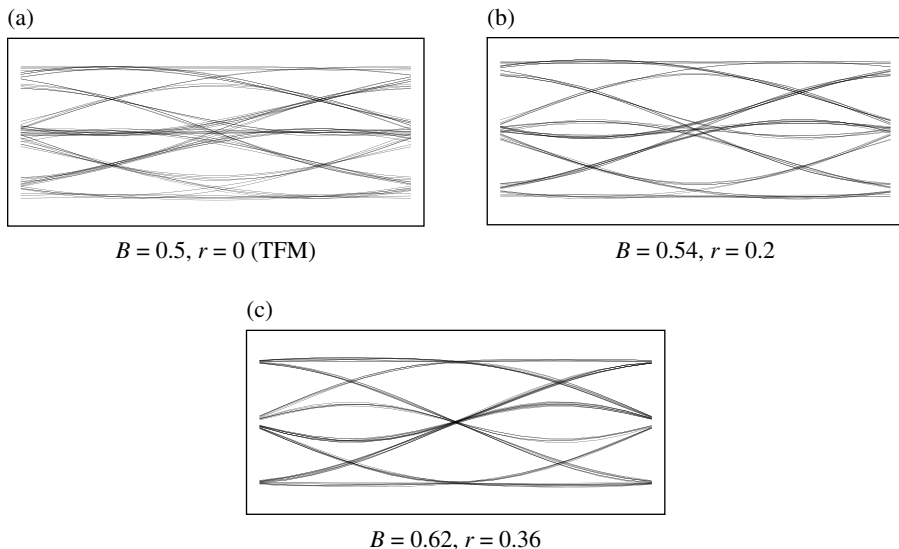
under the constraint that  $2A+B=1$ . TFM is obtained as the special case  $A=1/4$  and  $B=1/2$ . The transfer function of the partial-response filtering is given as

$$G(\omega) = B + 2A \cos(\omega T)\tag{6.16}$$

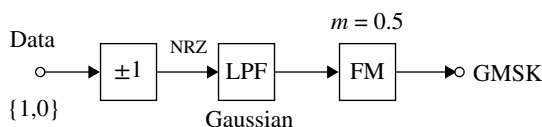
The Nyquist III filter is still employed.

Changing the parameters  $A$  and  $B$  and the roll-off factor for the Nyquist III filter, the spectrum and the bit error rate performance can be varied. For example, the eye diagram of the receiver baseband signal obtained with frequency discriminator detection is as shown in Figure 6.21. We can see the almost ideal three-level eye opening at the middle of successive symbol times for  $B=0.62$  and roll-off factor of 0.36 in Figure 6.21c. This is due to the fact that Nyquist's second criterion (Section 3.1.2) is satisfied by choosing those parameters. Three-level detection instead of five-level detection is adopted, since a three-level decision yields better error rate performance than a five-level decision. A rectifier and a two-level threshold detector can be employed for the three-level detection. This is because the three-level eye opening is produced from the interference between the successive symbols; that is, the effect is equivalent to the duobinary coding. The differential encoding for (generalized) TFM acts as pre-coding for the duobinary coding. Experimental results [19] with frequency





**FIGURE 6.21** Eye diagram of generalized tamed FM with frequency discriminator detection.

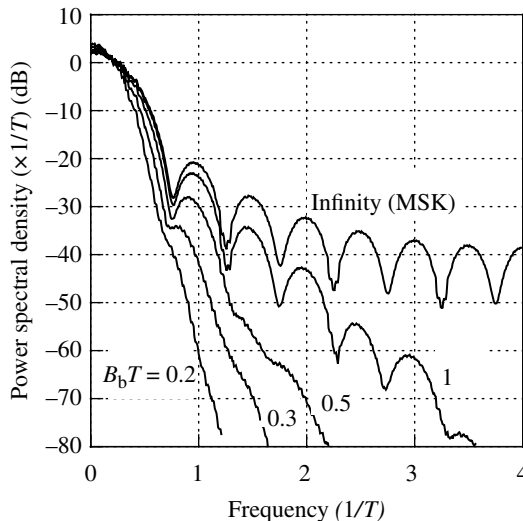


**FIGURE 6.22** GMSK signal generator.

discriminator detection of the generalized TFM show superior performance compared to coherent detection in fast-fading conditions. A maximum likelihood sequence estimation (MLSE) (Section 3.3.8) was applied to the three-level partial-response system with frequency discriminator detection. The error rate performance of frequency discriminator MLSE detection of GTFM in a fast-fading channel is reported in [20]. The modulation index does not need to be 0.5 as long as frequency discriminator detection is applied.

**6.2.2.4 Gaussian-filtered MSK** Gaussian-filtered MSK (GMSK) [4, 21, 22] was proposed by Hirade and Murota in 1979. The block diagram of the GMSK modulator is shown in Figure 6.22. GMSK modulation is performed by inserting a Gaussian low-pass filter as a premodulation filter to MSK modulation. When we assume an impulsive waveform, the transfer function of the premodulation is given as

$$H(\omega) = \frac{\sin(\omega T/2)}{\omega T/2} \exp \left[ -\ln \sqrt{2} \left( \frac{\omega}{2\pi B_b} \right)^2 \right] \quad (6.17)$$



**FIGURE 6.23** GMSK signal power spectra.

where  $B_b$  is the 3 dB bandwidth (Hz) of the Gaussian filter. The first term corresponds to an NRZ pulse waveform. The GMSK spectrum can be controlled by changing the bandwidth of the Gaussian filter (Fig. 6.23). The spectrum for  $B_b T = 0.21$  is almost equivalent to the spectrum of TFM.

A GMSK system can be seen as partial-response digital FM system in which the degree of intersymbol interference changes continuously with the Gaussian filter bandwidth ( $B_b$ ). The ability of GMSK to continuously control the band limitation of the premodulation signal is an advantage over other partial-response FM systems. GMSK never takes fixed phase points, since the premodulation filter does not satisfy the Nyquist III criterion.

Coherent, differential, and frequency discriminator detection can be applied to GMSK, as well as to MSK. Eye diagrams of GMSK with coherent detection are shown in Figure 6.24. The eye diagram is similar to that of TFM. As we increase  $B_b T$ , the eye diagram becomes close to the MSK eye diagram. Previously discussed demodulators for MSK can also be applied to GMSK.

Experimental results on bit error rates for coherent detection of GMSK with a Gaussian IF band-pass filter are shown in Figure 6.25 [4]. The optimum normalized 3 dB bandwidth  $B_b T$  for the Gaussian band-pass filter is reported as  $B_b T \approx 0.63$ . The degradation of the bit error rate performance in terms of  $E_b/N_0$  is 1.6 dB for  $B_b T = 0.25$  compared with the ideal antipodal transmission system.

Differential detection of GMSK is discussed in [23–30]. One-bit and two-bit delay detection are considered. The received eye patterns are shown in Figure 6.26. The eye opening becomes narrower due to intersymbol interference with a smaller  $B_b T$  value. Theoretical error rates for GMSK with one-bit and two-bit delay detection are described in [29]. Two-bit detection shows superior performance over one-bit detection. The optimum bandwidth ( $B_b T = 0.9$ – $1.4$  depending on  $B_b T$ , the detection scheme, and bit error rate) of the IF band-pass Gaussian filter is considerably wider

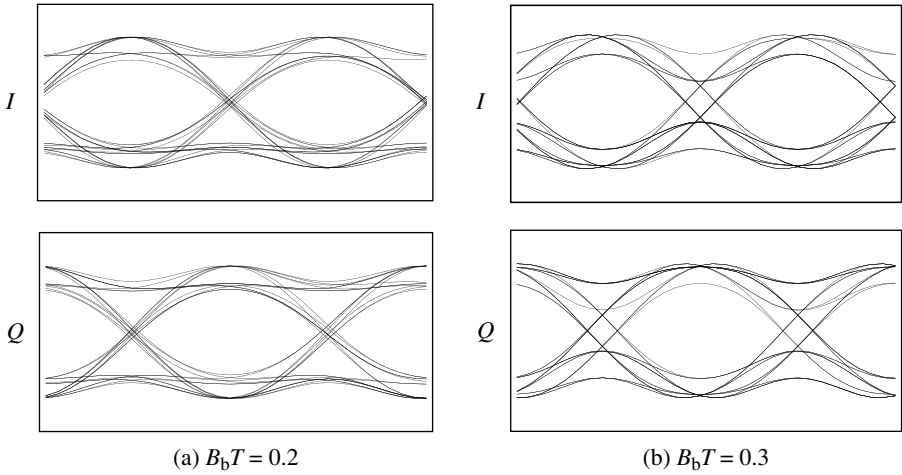


FIGURE 6.24 GMSK eye diagram with coherent detection.

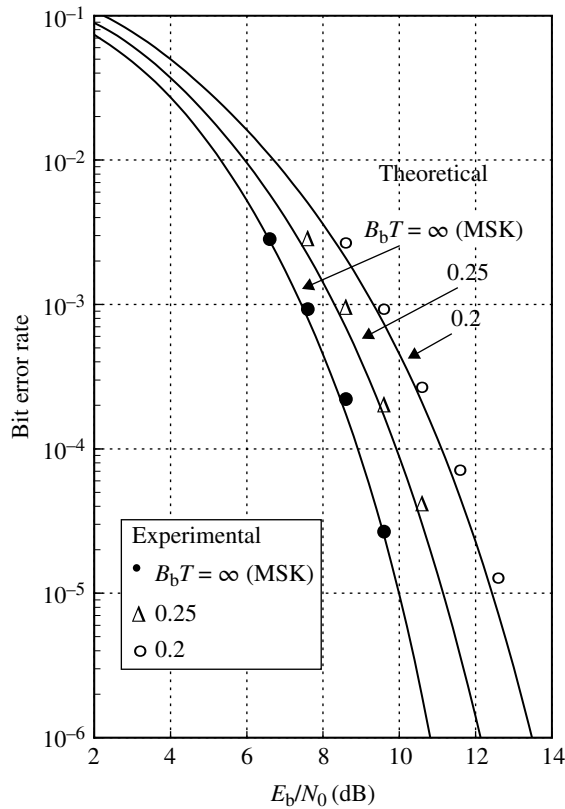
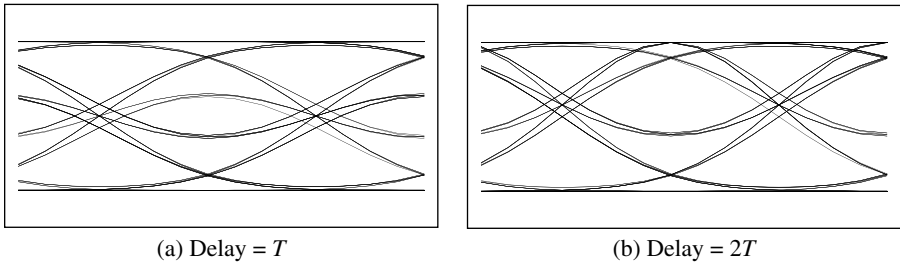
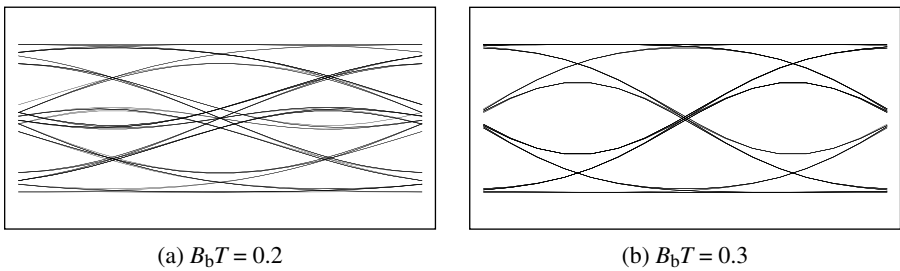


FIGURE 6.25 Experimental bit error rate for GMSK with coherent detection [4]. Reproduced by permission of Copyright © 1981, IEICE.



**FIGURE 6.26** GMSK eye diagram with differential detection ( $B_b T = 0.25$ ). Gaussian band-pass filter is used ( $B_{ff} T = 1.25$ ).



**FIGURE 6.27** GMSK eye diagram with frequency discriminator detection.

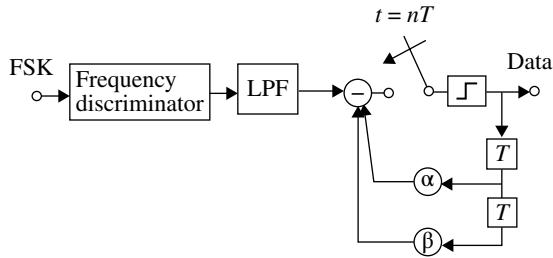
for differential detection than for coherent detection. Differential detection of GMSK using decision feedback is described in [30]. Owing to the improved eye opening with decision feedback, bit error rate performance is significantly improved.

Frequency discriminator detection of GMSK is discussed in [31–34]. With this detection, the modulation index can take an arbitrary value. The received eye diagrams are shown in Figure 6.27. In order to get wider eye openings, adaptive multi-level threshold decision was introduced [31] as shown in Figure 6.28. This scheme is equivalent to decision feedback detection (Section 7.3.3). Bit error rate performance is shown in Figure 6.29.

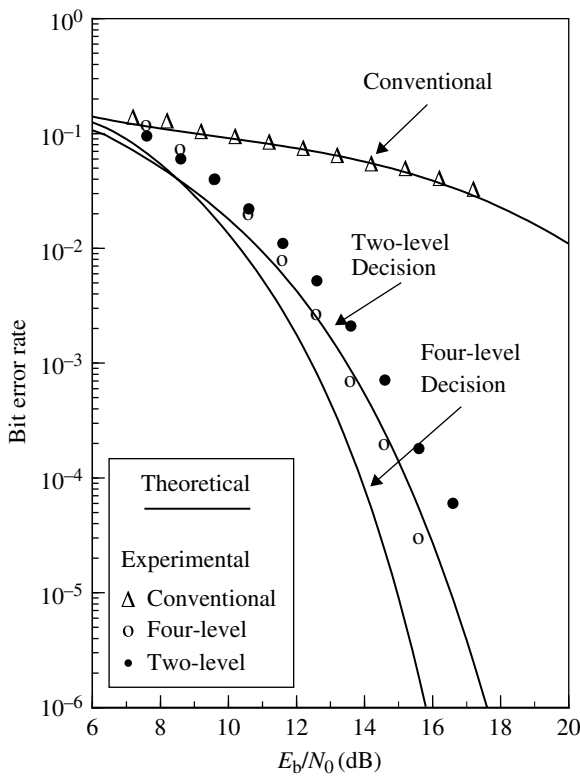
A three-level decision (Section 6.2.2) at the instant between successive symbol times was described in [32]. The error rate performance obtained with this method is almost the same as with the adaptive threshold multilevel decision. The error rate performance of differential and frequency discriminator detection of digital FM, including GMSK, is analyzed theoretically in [35].

The spectrum efficiency of a cellular system with GMSK modulation is analyzed in [36]. As a result of compromise between the spectrum bandwidth and cochannel interference,  $B_b T = 0.25$  and forward error correction with rate of 4/5 yields the maximum spectrum efficiency.

**6.2.2.5 Compact Spectrum Constant Envelope PSK (CCPSK)** This scheme was proposed in 1979 [37, 38]. CCPSK is a constant envelope continuous-phase PSK,

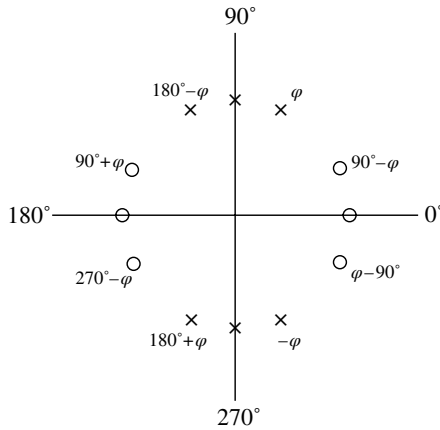


**FIGURE 6.28** Adaptive multilevel threshold decision for GMSK with frequency discriminator detection.



**FIGURE 6.29** Bit error rate performance of GMSK with frequency discriminator detection using adaptive threshold detection [31]. Reproduced by permission of Copyright © 1984, IEEE.

where phase transitions in symbol times are governed by 3-bit input sequences. CCPSK signals take one of 12 signal points marked by circles and crosses in turn as shown in Figure 6.30. The parameter  $\varphi$  takes an arbitrary value. If  $\varphi = 45^\circ$  or  $0^\circ$ , we have the phase constellations for TFM and MSK, respectively. Similarly, we have a two-level eye-opening for the inphase and quadrature-phase components. CCPSK



**FIGURE 6.30** Phase constellation for CCPSK. Phase transition occurs from an o-point marked to an x-point and vice versa.

is a generalized 12PM3 modulation method [39], where 12 stands for 12 phase points, PM means phase modulation, and 3 corresponds to the fact that the phase transition is governed by 3-bit sequences. Some waveforms for phase transition are suggested to maintain phase continuity. The derivative of the phase waveform, that is, the frequency waveform, is not necessarily continuous and causes poor out-of-band radiation for CCPSK.

**6.2.2.6 Correlative Phase Shift Keying** Correlative phase shift keying (CORPSK) is a special form of constant envelope PSK [18] that is defined by the following three points: (i) the information is encoded into phase shifts  $\Delta\varphi_m$ , so that the phase is continuous. Over one symbol interval, the phase shift is

$$\Delta\varphi_m = \varphi[(m+1)T_s] - \varphi(mT_s) = c_m \frac{2\pi}{n} \quad (6.18)$$

where the integer  $c_m$  carries the information and the constant  $n$  is the number of possible phase points. (ii) Successive phase points are correlated due to correlative encoding of input data, which takes  $L$  ( $\geq 2$ ) levels; nonlinear and linear correlative encodings are considered. (iii) The phase waveform, as well as its time derivative (frequency waveform), is continuous.

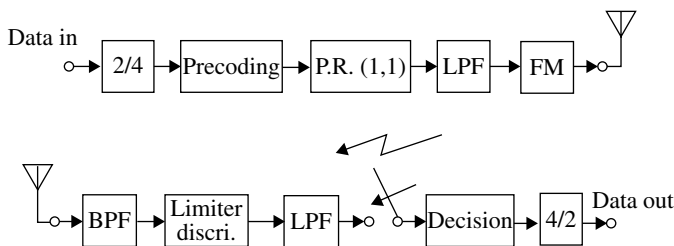
Thus, CORPSK includes many varieties of constant envelope PSK with correlative encoding. A new method of CORPSK is the four-level input, nonlinear correlative encoding system (CORPSK (4-5)). In this system, phase points are the same as in QPSK. The phase transition of  $\pi$  or  $-\pi$ , both of which result in the same phase point, is nonlinearly controlled so that the phase path takes a smoother trajectory. CORPSK shows almost the same out-of-band radiation as TFM but has better bit error rate performance.

**6.2.2.7 Digitally Phase-Modulation (DPM) Signals** The term, DPM is used for constant envelope PSK with band-limited digital input signals. Digital-phase modulation with correlative coding was analyzed by Maseng [40]. In this system, modulated signals never take fixed points, unlike in phase shift keying systems. Hence, symbol-by-symbol phase detection on the inphase and quadrature-phase plane is not applicable. Coherent demodulation with MLSE is used for the receiver. A residual carrier component, which appears in the digitally phase-modulated signal, is an attractive feature of the DPM signal for coherent demodulation. Modems for DPM are easier to implement than for digital FM, since the phase of a DPM signal never depends on the preceding data, in contrast to digital FM. Compared to binary phase shift keying, a coding gain of several decibels is obtained with a higher modulation index.

**6.2.2.8 Continuous-Phase Modulation** Continuous-phase modulation (CPM) has a constant envelope continuous-phase (digital) waveform. Multilevel modulation, various modulation indices, and premodulation waveforms including partial-response signaling are considered. The optimum (sequence) detector with coherent demodulation is assumed. The demodulated signal at a symbol period is correlated with those at previous periods due to an arbitrary modulation index. Even when the premodulation waveforms have no correlation between different symbol periods (full response system), the correlation operation is still performed. Therefore, the optimum detector observes the received signal for more than one symbol interval.

Spectrum and bit error rate performance are investigated extensively in [41–43]. Some CPM signals show superior performance compared to MSK and PSK signals. Multi- $h$  modulation, where the modulation index is periodically varied, was proposed to increase distance between signal sequences [44].

**6.2.2.9 Duoquaternary FM** If we apply the four-level signal to a partial-response filter whose impulse response is  $h(t) = \delta(t) + \delta(t - T)$  ( $T$ : symbol duration), we have the duoquaternary signal. Duoquaternary FM with limiter–frequency discriminator detection (see Fig. 6.31) was discussed in [45, 46]. A seven-level signal is obtained at the output of the postdemodulation filter. The spectrum and bit error rate performance are better than four-level FM.



**FIGURE 6.31** Duoquaternary FM with frequency discriminator detection.

### 6.2.3 Nyquist-Filtered Digital FM

**6.2.3.1 Nyquist-Filtered Multilevel FM** In 1979, Akaiwa, Takase, Ikoma, and Saegusa [47, 48] proposed digital FM with frequency discriminator detection, where the premodulation and postmodulation low-pass filters as a whole meet Nyquist's first criterion (see Fig. 6.32). When the bandwidth of the band-pass IF filter is wide enough, intersymbol interference-free transmission is achieved. The premodulation filter, together with a low modulation index and multilevel signaling, contributes to a narrow spectrum of the modulated signal. The postdemodulation filter helps achieve good bit error rate performance by band limiting the demodulated noise.

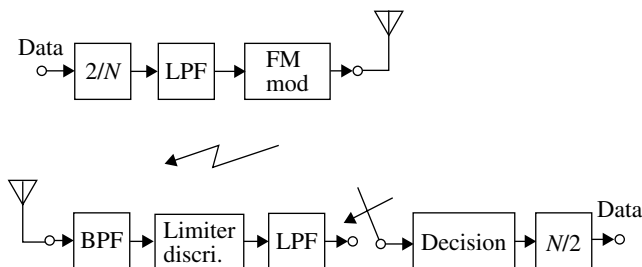
The spectrum and bit error rate performance are shown in Figures 6.33a, b and 6.34a, b. The Nyquist I filter characteristics are equally divided into the premodulation and postdemodulation filter. A sharp cutoff band-pass filter with the Carson bandwidth (Section 5.3.2) is assumed.

The effects of the IF band-pass filter bandwidth on received eye diagrams and bit error rate are shown in Figures 6.35 and 6.36, respectively. The crosses in Figure 6.36 denote computer-simulated results, and the solid lines show theoretical results described in Section 5.5.4.

The bit error rates for Nyquist-filtered four-level FM under Rayleigh fading are shown in Figure 6.37. Solid lines show theoretical results obtained by adding the error rates given by Equations 5.53 and 5.56 without diversity reception.

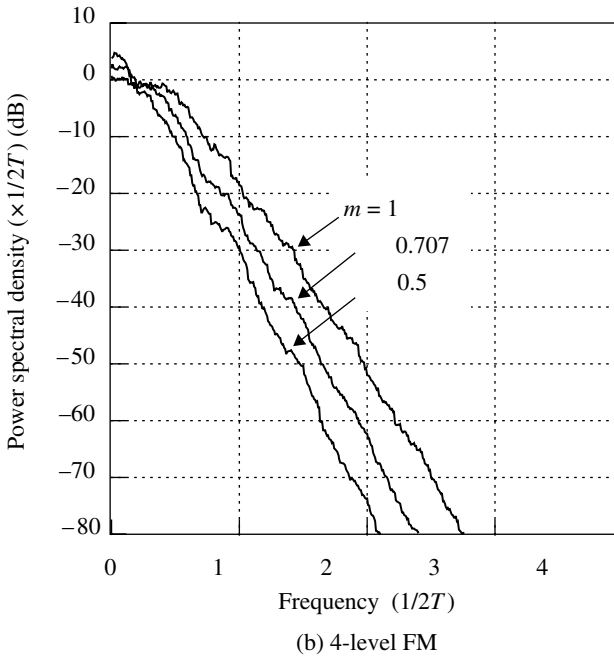
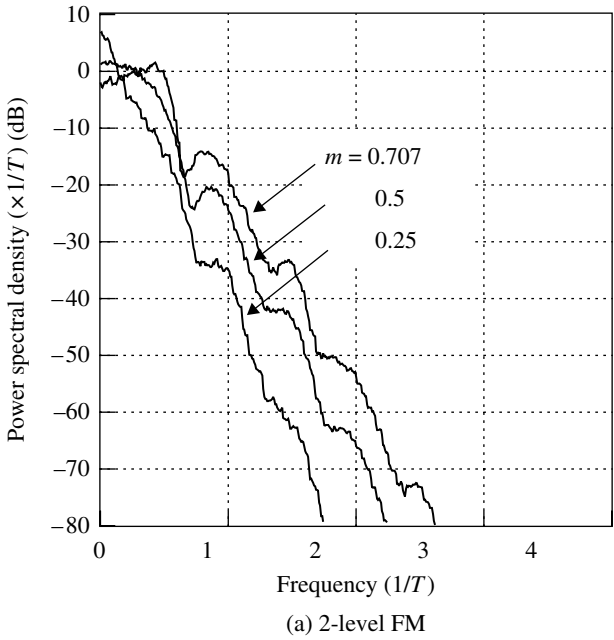
If we adopt the Nyquist third characteristics (Section 3.1.3) to the premodulation filter and choose a proper value for the modulation index, we have constant envelope phase shift keyed signals: for two-level and four-level signaling systems, constant envelope  $\pi/2$ -shifted BPSK and  $\pi/4$ -shifted QPSK signals are obtained with modulation index of 0.5 and 0.75, respectively [49]. Since these signals take fixed signal phases, coherent and differential detection can be used to demodulate the signal. Spectrum and bit error rate performance of the four-level system are shown in Figures 6.38 and 6.39, respectively, and are almost equal to the performance of TFM.

**6.2.3.2 PLL-QPSK** PLL-QPSK was proposed in 1979 [50, 51]. The block diagram of PLL-QPSK is shown in Figure 6.40.  $\pi/4$ -shifted QPSK with an NRZ pulse shape is fed to a phase-locked loop (PLL) circuit to produce a constant envelope narrow-band signal. A low-pass characteristic is given to the loop transfer function to suppress out-of-band radiation.

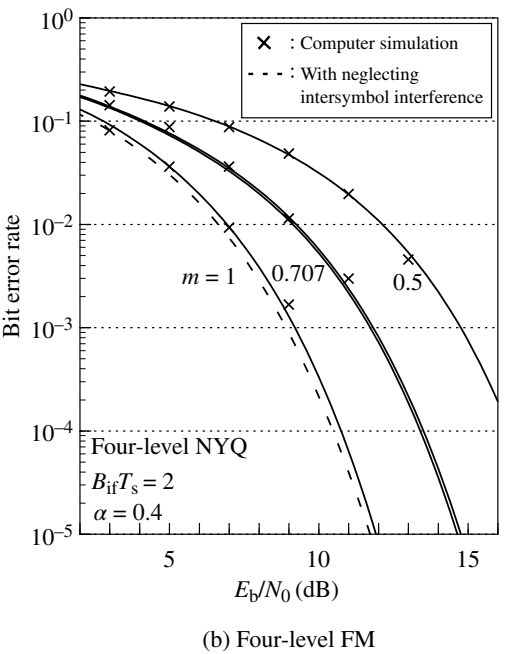
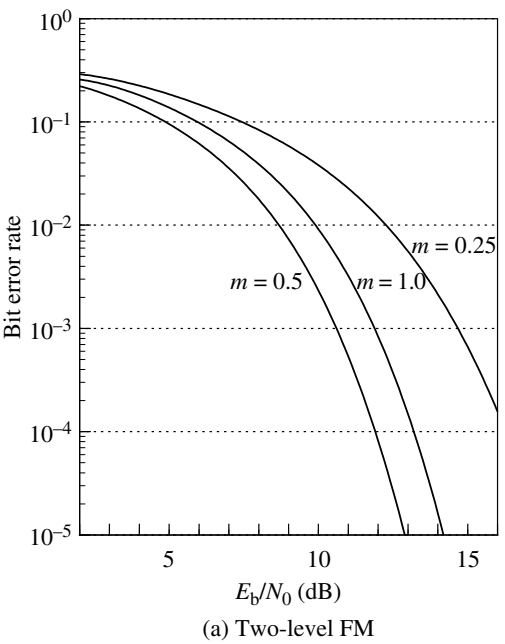


**FIGURE 6.32** Block diagram of  $N$ -level digital FM with frequency discriminator detection.

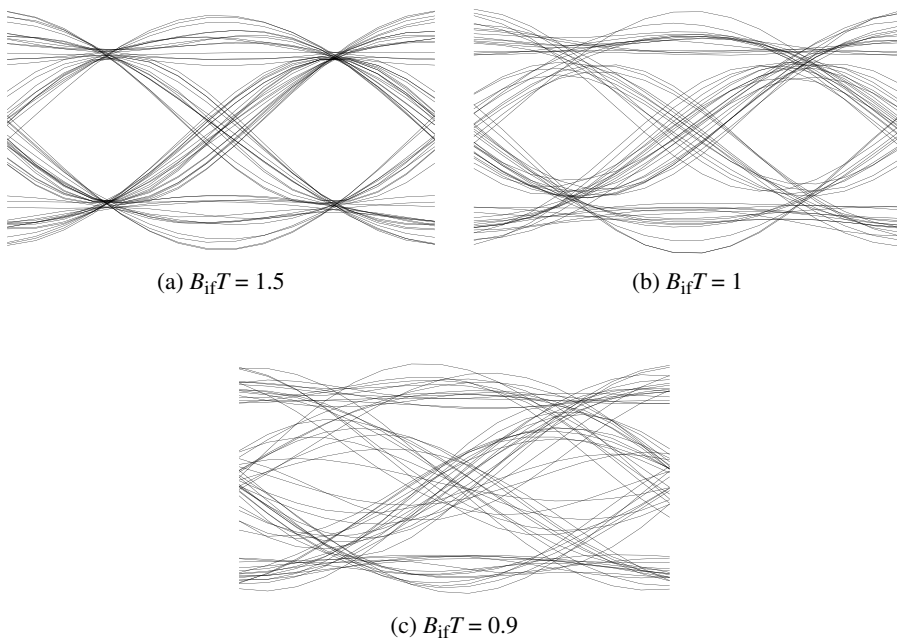




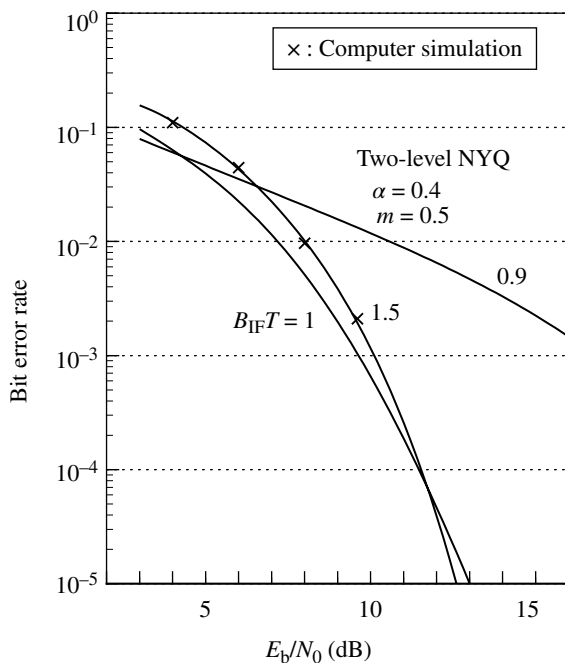
**FIGURE 6.33** Spectra of square root of Nyquist I-filtered digital FM with modulation index as a parameter. The roll-off factor is 0.5.  $T$  is the bit duration.



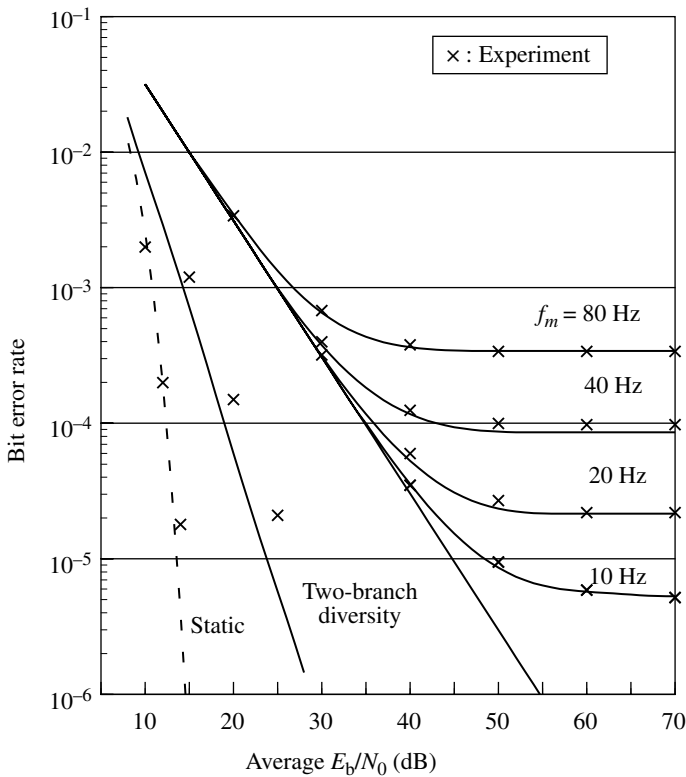
**FIGURE 6.34** Bit error rate for the Nyquist-filtered digital FM with modulation index as a parameter.



**FIGURE 6.35** Effects of IF band-pass filter bandwidth on received eye diagrams for Nyquist-filtered two-level FM. The modulation index = 0.5, the roll-off factor is 0.4, and the band-pass filter characteristics is its rectangular shape.



**FIGURE 6.36** Effects of IF band-pass filter bandwidth on bit error rate for Nyquist-filtered two-level FM.

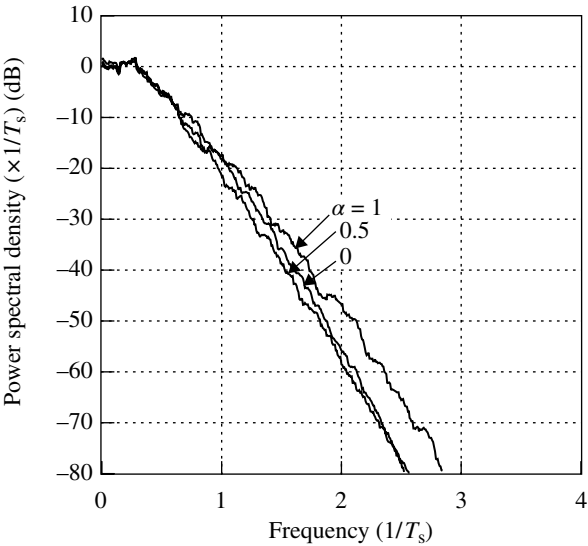


**FIGURE 6.37** Bit error rates for Nyquist-filtered four-level FM under Rayleigh fading. In plot,  $f_m$  denotes the maximum Doppler frequency.

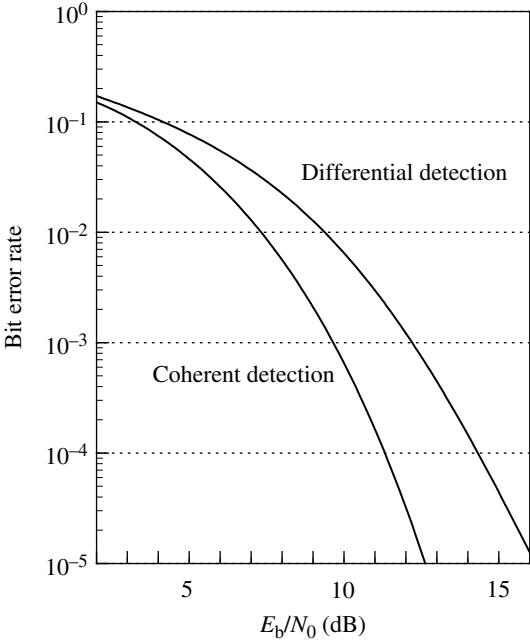
A limiter–frequency discriminator detector with an I&D filter is employed. It detects the phase shifts of  $\pm\pi/4$  and  $\pm3\pi/4$  of the  $\pi/4$ -shifted QPSK signal. Therefore, differential encoding at the transmitter is assumed. The differential encoding of the input data signal can be performed in FM modulation through its integrating process. Thus, the PLL-QPSK signal can be produced by FM modulation as shown in Figure 6.41d. Consequently, we can see from Figure 6.41 that the PLL-QPSK is a special case of four-level FM, where the modulation index is 0.75. The equivalent premodulation filter for the FM modulator and the postdemodulation filter including the I&D filter, as a whole, satisfy Nyquist’s first criterion.

#### 6.2.4 Performance Comparison

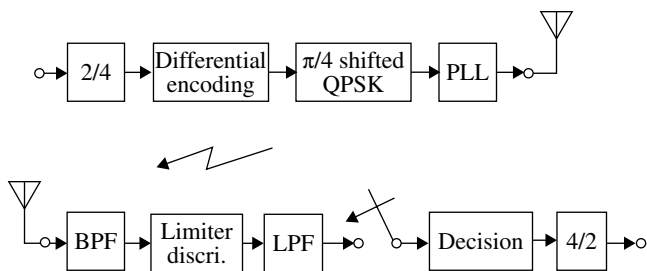
A simultaneous comparison of spectrum and bit error rate performance between digital FMs is shown in Figure 6.42 [52]. The spectrum performance is represented by the attenuation of the power spectral density at a frequency that deviated by bit rate frequency from the center frequency. For the power spectral density attenuation of around  $-60$  dB, the duoquaternary system shows the best performance. The actual



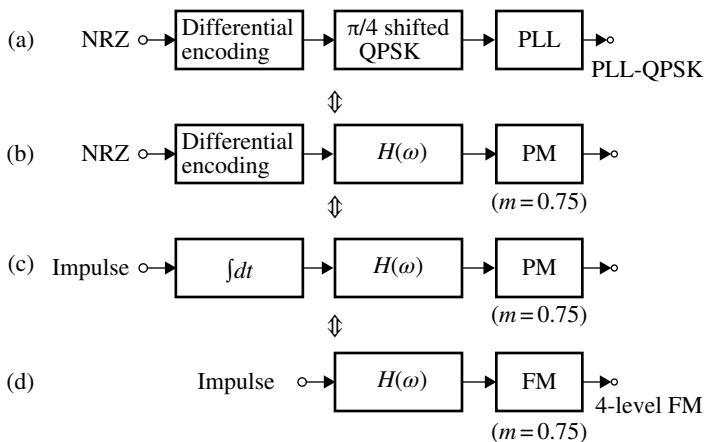
**FIGURE 6.38** Spectra of Nyquist III-filtered four-level FM with a modulation index of 0.75 with roll-off factor as a parameter.



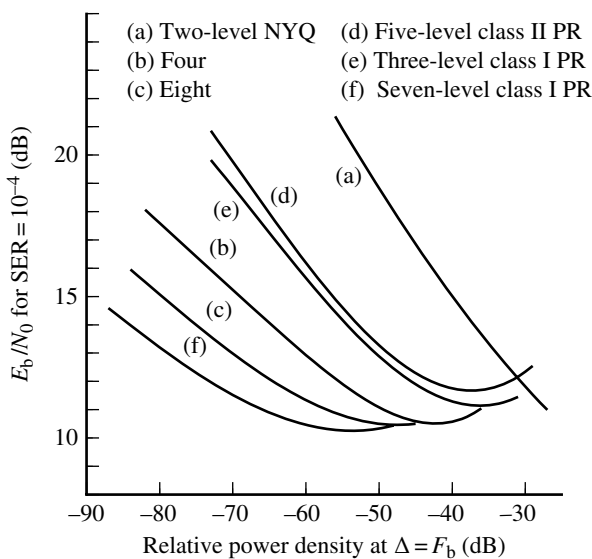
**FIGURE 6.39** Bit error rates for Nyquist III-filtered four-level FM with coherent and differential detection. Differential coding is assumed. The modulation index is 0.75 and the roll-off factor is 0.5. The band-pass filter has a rectangular transfer function with a normalized bandwidth of  $0.8/T$  ( $T$ = bit duration).



**FIGURE 6.40** Block diagram of PLL-QPSK.



**FIGURE 6.41** Equivalence between PLL-QPSK and four-level FM.



**FIGURE 6.42** Performance comparison between digital FMs with limiter-frequency discriminator detection. The curves are obtained by changing modulation index.

choice from those systems must take into consideration the circuit complexity and performance degradation due to imperfections in the circuit.

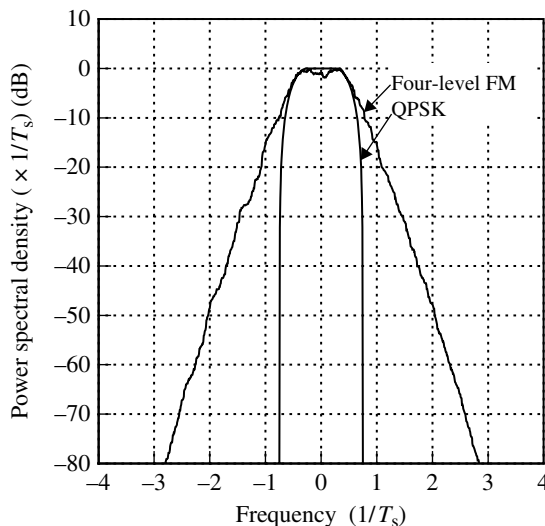
### 6.3 LINEAR MODULATION

The spectrum of linearly modulated signals is given by translating the modulating baseband signal spectrum into the carrier frequency (Section 5.4.1). Thus, filtered linear modulation yields a narrow spectrum corresponding to the filtering at the baseband frequency. A penalty for the narrow spectrum is that the digital linear modulation loses its constant envelope property as with unfiltered signaling.

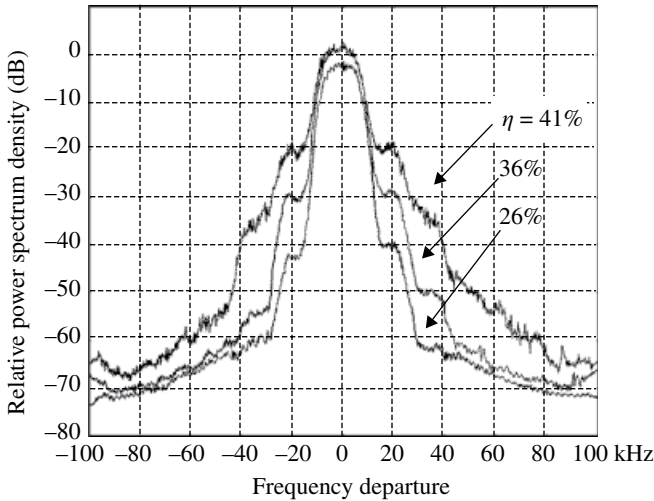
Figure 6.43 compares the spectra for filtered linear QPSK and premodulation filtered four-level FM. The superiority of the linear modulation over the constant envelope (nonlinear) modulation is clear: two to three times narrower bandwidth at an out-of-band radiation of around  $-60$  dB, for example, is obtained with the linear modulation.

However, this is true only in an ideal channel. If the transmitter shows nonlinear distortion, the linear modulation loses the spectral superiority, since the out-of-band radiation appears proportionally to the degree of odd order nonlinear distortion (Appendix 5.A). For example, Figure 6.44 shows spectra for  $\pi/4$ -shifted QPSK at the output of a quasilinear transmit power amplifier. The out-of-band radiation increases with the increase in the power efficiency of the amplifier: a higher power efficiency is achieved with a higher nonlinear distortion in the power amplifier. The power efficiency is a prime issue in mobile radio communication, especially for portable communication, where batteries are the power sources.

Suppressing the out-of-band radiation after the power amplifier by the use of a band-pass filter is impractical in mobile communication since the center frequency



**FIGURE 6.43** Spectra for QPSK and four-level FM with modulation index of 0.5. Square root of Nyquist I characteristic with a roll-off factor of 0.5 is assumed.



**FIGURE 6.44** Spectra of  $\pi/4$ -shifted QPSK at the output of a 900 MHz MOSFET quasilinear power amplifier for different input power ( $\eta$  = power efficiency). Roll-off factor of 0.5 is used.

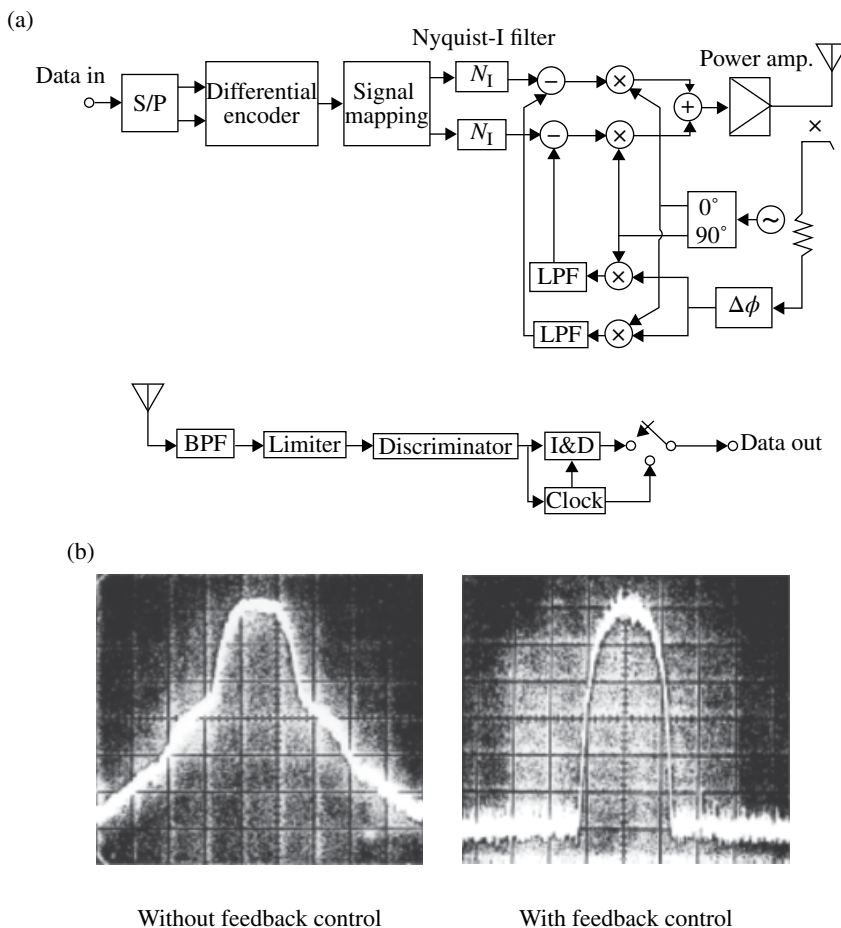
must be varied. Even when the center frequency is fixed, the realization of a low-loss band-pass filter is difficult, since the ratio of passband bandwidth to the center frequency is very small in typical mobile radio communications.

It is difficult to get a linear power amplifier that is power efficient and at the same time shows low nonlinear distortion. This was one reason why linear modulation was discarded for application to mobile radio communication [53]. Another difficulty with the linear modulation for mobile radio communication was the implementation of the receiver circuit: in order to receive a linearly modulated signal, an automatic gain control (AGC) circuit was required that could cope with fast-fading and wide dynamic range of the received signal in mobile radio channels. Such an AGC circuit is much harder to implement than a limiter circuit, which can be used for receiving a constant envelope signal. This is especially so for narrow-band signal transmission.

In 1985, Akaiwa and Nagata proposed a linear  $\pi/4$ -shifted QPSK modulation system (Fig. 6.45a) that resolved the abovementioned difficulties [54, 55]. A feedback control was introduced to compensate for the nonlinear distortion of a power-efficient amplifier and a limiter circuit was adopted in the receiver. The spectral performance is improved by 30 dB, as shown in Figure 6.45b, owing to the negative feedback control (Example 2.11). Out-of-band radiation of -60 dB was achieved with a power efficiency of 35%. Bit error rate performance was comparable with performance for digital FM. Since then, many reports have been published on linear modulation systems for digital mobile communications [56–59]. In addition to a narrower signal spectrum, a linear modulation system employing linear circuits at the transmitter shows advantages such as ease of transmitter output power control and tailoring of the on-off burst signal to prevent spectrum spreading (SS) for TDMA signals.

In order for digital linear modulations to show a superior spectrum compared to narrow-band constant envelope modulations such as TFM, PLL-QPSK, and Nyquist-filtered





**FIGURE 6.45** (a) Block diagram of linear  $\pi/4$ -shifted QPSK system with a linearized power amplifier and limiter–discriminator detection. (b) Spectrum improvement with negative feedback control. Vertical axis, 10 dB/division; horizontal axis, 10 kHz/division.

multilevel FM, multilevel modulation must be introduced. The candidates are QPSK and its modifications OQPSK and  $\pi/4$ -shifted QPSK, 8PSK, 16QAM, and other higher-level modulations.

This section describes linear modulation/demodulation systems. Nonlinearity compensating techniques required for linear modulation systems are described in Section 8.6.3.

### 6.3.1 $\pi/4$ -Shifted QPSK

The same spectra are given by QPSK, OQPSK, and  $\pi/4$ -shifted QPSK. Their trajectories are shown in Figure 6.46. OQPSK and  $\pi/4$ -shifted QPSK never pass through the origin, where the amplitude becomes zero. This property has advantages when

the signals are applied to an amplitude limiter together with noise at a receiver: the phase perturbation due to noise becomes less. Furthermore, a lower dynamic range of the signal amplitude for OQPSK and  $\pi/4$ -shifted QPSK is beneficial from the viewpoint of transmitter and receiver circuit implementation. In order to express the dynamic range of signal level, the complementary cumulative distribution function (CCDF:  $G(x)=1-F(x)$ ) of signal instantaneous power normalized with average power is shown in Figure 6.47. A lower instantaneous power normalized with average power is shown in the order of OQPSK,  $\pi/4$ -shifted QPSK and QPSK.

Coherent detection can be applied to all the three modulation methods. Differential detection cannot be applied to OQPSK, since the signal phase never takes fixed points, as shown in Figure 6.48, and consequently the demodulated signal eye never opens.

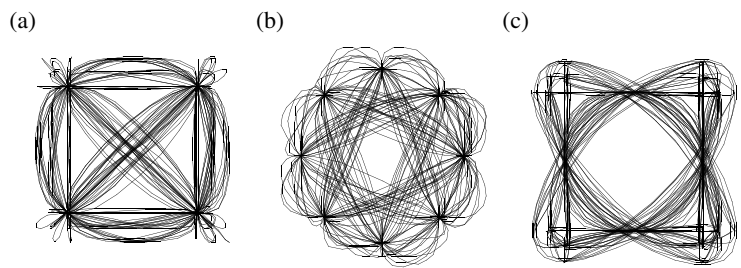


FIGURE 6.46 QPSK signal trajectories. (a) QPSK, (b)  $\pi/4$ -shifted QPSK, and (c) offset QPSK.

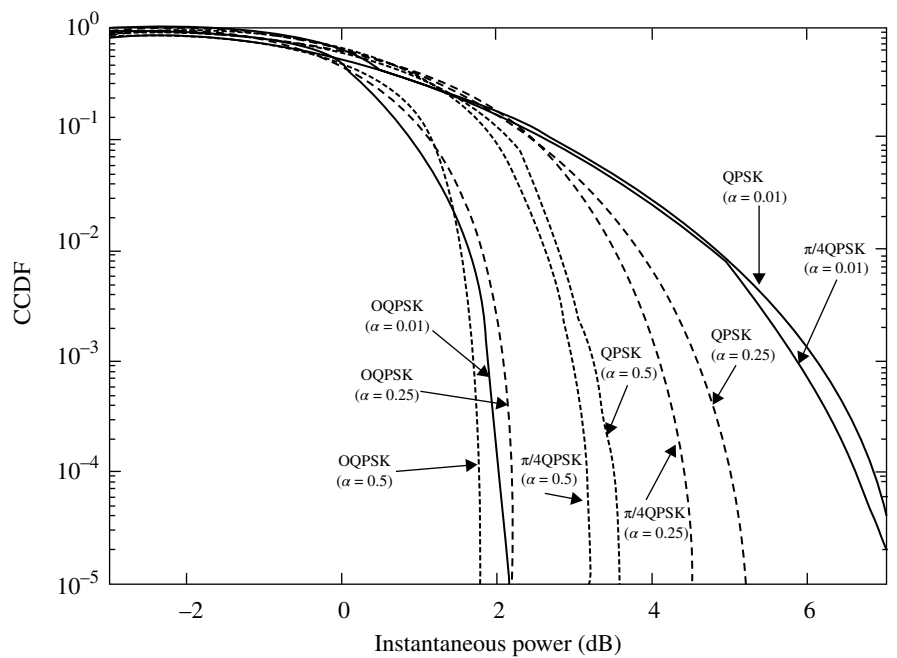
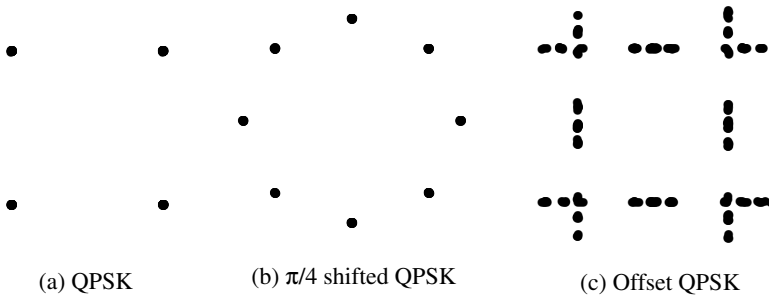
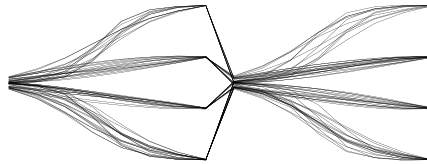


FIGURE 6.47 Complementary cumulative distribution function of normalized signal power.



**FIGURE 6.48** Signal phase points at sampling instants.



**FIGURE 6.49** Eye diagram of limiter-frequency discriminator and integrate-and-dump detection of  $\pi/4$ -shifted QPSK Nyquist I filter with roll-off factor of 0.5 is assumed.

Based on the aforementioned consideration,  $\pi/4$ -shifted QPSK was selected and investigated for application to digital mobile radio communication in [55]. A limiter-frequency discriminator with I&D detection was proposed (Fig. 6.45a). The principle of this system is to detect phase shifts of  $\pm\pi/4$  and  $\pm3\pi/4$  associated with the  $\pi/4$ -shifted QPSK signal, as explained mathematically in the following.

The output of the frequency discriminator is proportional to the instantaneous frequency  $\omega_i(t)$ . At the output of the I&D filter, we have

$$s_d(nT+t) = \int_{nT}^{nT+t} \omega_i(\tau) d\tau \quad (0 \leq t \leq T) \quad (6.19)$$

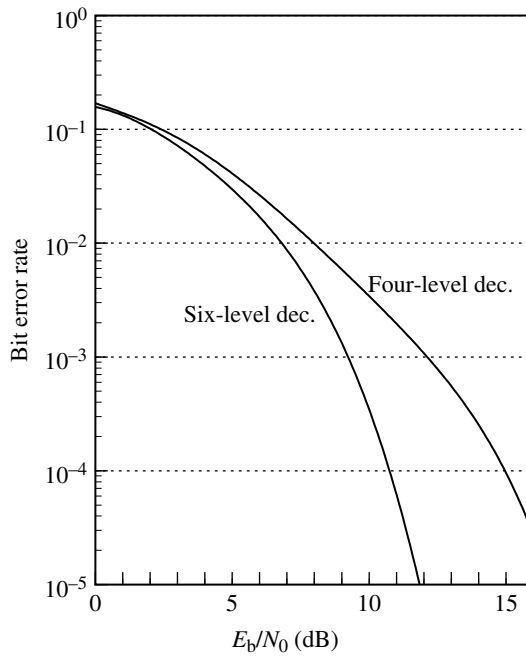
where  $T$  is the symbol duration. At the sampling instant  $(n+1)T$ , we have

$$s_d(nT+T) = \Delta\theta(nT) = \theta(nT+T) - \theta(nT) \quad (6.20)$$

where the phase  $\theta(t)$  is given by  $\theta(t) = \int_{-\infty}^t \omega_i(\tau) d\tau$  and  $\Delta\theta(nT)$  represents the phase shifts during one symbol duration.

Differential encoding is assumed in the transmitter. The received signal eye diagram is shown in Figure 6.49. Raised-cosine roll-off Nyquist I filtering is employed at the transmitter. Intersymbol interference is caused by the sharp cutoff band-pass filter at the receiver.

The bit error rate performance is shown in Figure 6.50. With the aforementioned detection system, errors occur more often for the phase shift of  $\pm3\pi/4$  than for  $\pm\pi/4$ ,



**FIGURE 6.50** Bit error rate performance for  $\pi/4$ -shifted QPSK with limiter–frequency discriminator and integrate-and-dump filter detection. Raised-cosine roll-off Nyquist I filter with roll-off factor of 0.5 is assumed at the transmitter. Gray code mapping is assumed.

since the signal trajectory passes closer to the origin. An error occurs when the phase trajectory is deviated by noise such that the resultant signal's phase rotation may take a direction opposite to its original path. This results in a total phase shift of  $\pm 5\pi/4$  instead of  $\mp 3\pi/4$ . To prevent this kind of error, the phase shift of  $5\pi/4$  and  $-5\pi/4$  is corrected as  $-3\pi/4$  and  $3\pi/4$ , respectively (i.e., mod  $2\pi$  decision). This is done by making six-level instead of four-level decisions at the output of the I&D filter. A more general decision is made by applying the mod  $2\pi$  operation also to the phase shift of  $\pm\pi/4$  (i.e., eight-level decision). The bit error rate performance is improved with this technique, as shown in Figure 6.50.

Differential detection is also employed to detect the phase shift of  $\pm\pi/4$  and  $\pm 3\pi/4$ . The output of the differential detector at the sampling instants becomes

$$\begin{aligned}s_x(nT + T) &= \sin \Delta\theta(nT), \\ s_y(nT + T) &= \cos \Delta\theta(nT)\end{aligned}$$

The mod  $2\pi$  operation on  $\Delta\theta(nT)$  described earlier is automatically built in to the sine and cosine functions.

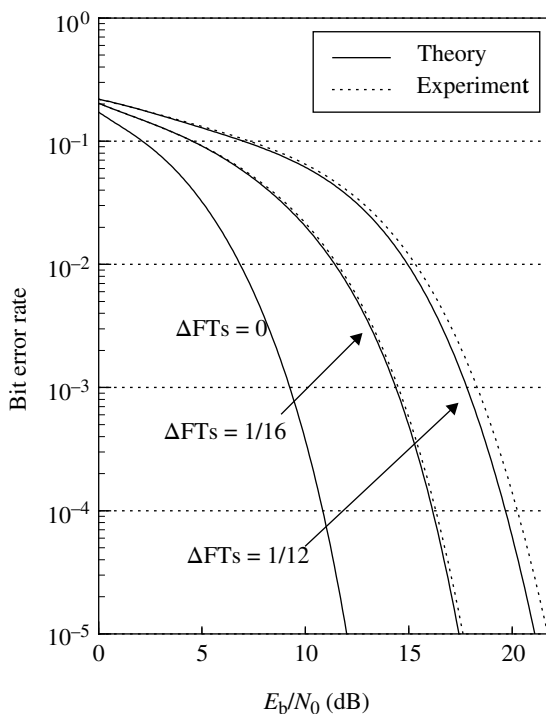
Since the four-level decision on the  $\Delta\theta(nT)$  is equivalent to the two-level decision on  $\sin[\Delta\theta(nT)]$  and  $\cos[\Delta\theta(nT)]$ , the same bit error rate performance is obtained with frequency discriminator detection with mod  $2\pi$  operation and I&D filter or differential

detector. This is true as long as the effect of a postdetection filter for the differential detector is neglected. This assumption is usually valid since the postdetection filter is only used to remove the higher-order IF frequency components produced in the multiplication circuit. In this case, the limiter does not affect the error rate performance.

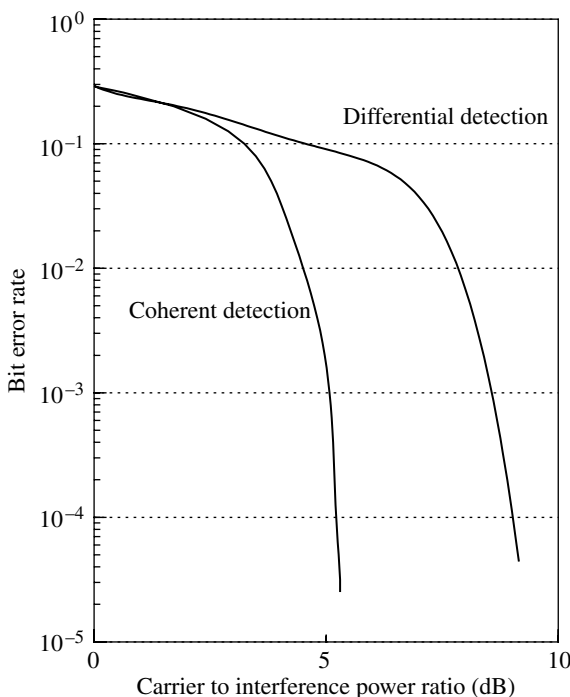
Equal splitting of the raised-cosine roll-off Nyquist I filter characteristics between the transmitter and receiver gives the optimum error rate performance for differential as well as coherent detection under additive white noise conditions. For differential detection, the receiver filter must be placed before the differential detector. A band-pass filter whose low-pass equivalent transfer function is the square root of the raised-cosine roll-off Nyquist I filter meets the requirements [57].

The bit error rate performance of ( $\pi/4$ -shifted) QPSK using differential detection and a matched band-pass filter is investigated in the following. Experimental results are given by computer simulations.

Figure 6.51 shows bit error rates versus bit energy to noise power ratio. The theoretical results are given by Equation 5.37. In the same figure, the error rates with the center frequency offsets are shown. Theoretical values are calculated with Equation 5.38. The discrepancy between the theoretical and experimental results is due to the fact that the intersymbol interference caused by the frequency offset is not



**FIGURE 6.51** Bit error rate performance for  $\pi/4$ -shifted QPSK with differential detection with frequency offset as a parameter.



**FIGURE 6.52** Bit error rate performance of  $\pi/4$ -shifted QPSK under cochannel interference environment.

considered in the theory. The experimental bit error rate performance under the cochannel interference environment is shown in Figure 6.52.

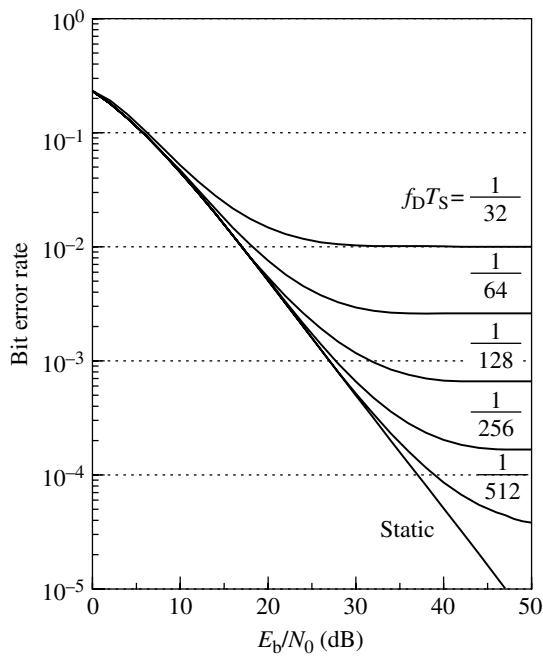
Under the Rayleigh fading condition, Figures 6.53 and 6.54 show the bit error rate performance for average bit energy to noise power ratio and cochannel interference, respectively.

### 6.3.2 Eight-Level PSK

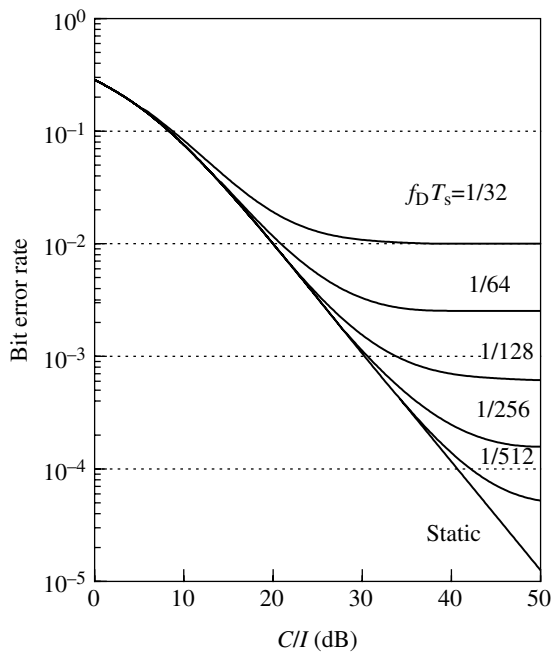
Eight-level PSK has a signal spectrum that is two-thirds of a QPSK signal spectrum. It may be the highest level modulation that falls into the category of phase shift keying and is adopted for actual use in mobile radio communication: higher-level modulation is probably adopted from QAM such as 16QAM.

If we combine a rate 2/3 error correcting code and 8PSK, the bandwidth of the modulated signal is the same as for QPSK. The error rate performance is improved with the error correcting code. 8PSK was investigated for application to mobile radio communications in [60].

8PSK is used for EDGE System (an evolution system of GSM; Section 9.7.4).



**FIGURE 6.53** Bit error rate performance of  $\pi/4$ -shifted QPSK with differential detection under Rayleigh fading condition.



**FIGURE 6.54** Error rate for  $\pi/4$ -shifted QPSK with differential detection in the presence of cochannel interference and Rayleigh fading condition.



P: Pilot signal

FIGURE 6.55 Insertion of pilot signal in the time domain.

6.3.3 16QAM

16QAM is a bandwidth-efficient modulation method. Because of the encoding of 4bits/symbol, 16QAM occupies half the bandwidth of QPSK. Because the signal constellation is spread in both amplitude and phase, noncoherent demodulation cannot be applied, which is a disadvantage in mobile radio communications. To make 16QAM applicable to mobile radio communications, coherent demodulation methods are proposed [61, 62] that are sufficiently effective under fast-fading conditions. The principal idea is to send a pilot (carrier) inserted in the modulated signal for stable operation of coherent demodulation. The extracted pilot signal is used to compensate for the perturbed amplitude and phase of the received signal passing through the fading channel. Two approaches are known for insertion of the carrier signal: one is in the frequency domain [61] and the other is in the time domain [62]. (For a BPSK system, there is another class of the carrier signal insertion [63], viz., insertion in the orthogonal component of the BPSK signal.)

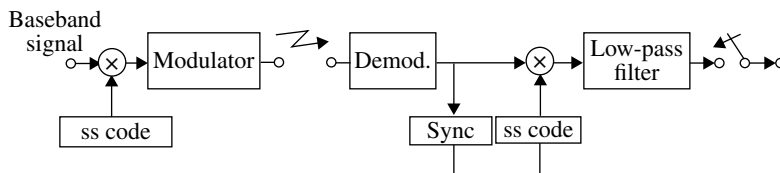
In the transparent tone-in-band (TTIB) method, the modulated signal is divided into two spectrum components in the frequency domain and they are shifted up and down to give a band gap into which to insert the pilot signal. At the receivers, the inverse process is carried out. The dual-pilot tone can be inserted at the edge of the modulated signal spectrum [64]. This technique is potentially effective for compensating for frequency-selective fading. Using the pilot signals, estimation of the frequency-selective transfer function over the signal bandwidth is possible.

The method of inserting the pilot signal in the time domain was proposed by Sampei [62]. A pilot symbol is inserted into each of  $N - 1$  symbols (Fig. 6.55). At the receiver, the pilot signal is extracted and interpolated to produce a time-continuous carrier signal. This method requires less signal processing than others. The interval of insertion of the pilot signal should be short enough to cope with fading speeds. For a transmission speed of 64 kbps, coherent detection is stable with pilot signal insertion in every 16-symbol and a second-order interpolator. Field trial experiment results are described in [65]. 64QAM that shows two times higher spectrum efficiency is standardized in the LTE system (Section 9.7.6).

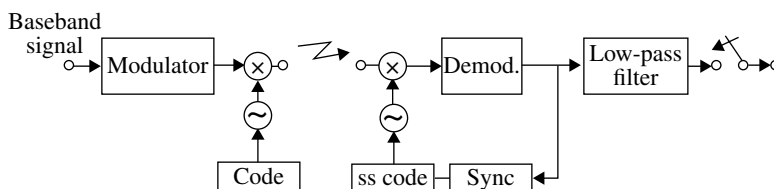
6.4 SPREAD-SPECTRUM SYSTEM

In spread-spectrum systems, the spectrum is spread by introducing additional modulation with SS codes. If FSK is used for the additional modulation, the spread-spectrum system is called a frequency hopping (FH), and if the baseband signal or





(a) Direct sequence system



(b) Frequency hopping system

**FIGURE 6.56** Spread-spectrum communication system.

the modulated signal is multiplied with an SS code, it is called a direct sequence (DS) system (Fig. 6.56a and b).

An SS code is a sequence consisting of so-called chips. Pseudo noise (PN) sequences are used for the SS codes, whose characteristics vary depending on the purpose of the system. With multiplication of a PN (SS) code to data signal, the spectrum spreads according to the spectrum bandwidth of the SS code: multiplying the PN code in the time domain results in convolution integral in the frequency domain (Eq. 2.40). If we use an SS code with a length of  $N$  chips for each data symbol, then the chip rate of the SS code is  $N$  times the data rate. As a result, the spectral bandwidth is increased (spread) by  $N$  times the original data spectral bandwidth. The spectrum for an FH system depends on the range of FH. If the hopping rate is slow (slow FH) compared to the data rate, its spectrum is given as a set of the modulated signal spectra for different (hopped) carrier frequency. If the hopping rate is faster (fast FH) than the data rate, its spectrum further spreads by an amount equal to a stationary signal spectrum multiplied with the number of hops per data symbol period.

The receiver for an SS system performs as follows. Let us consider first a DS system in Figure 6.56a. The demodulator outputs the data signal multiplied by the SS code. Assume that the SS code is a binary polar code (Section 3.2.1). Then by multiplying the demodulated signal with the SS code in synchronism to the received signal (despreading process), we have a narrow-band signal, which is the transmit baseband signal. Hence, this system can be seen as an ordinary system where the pulse waveforms take the SS codes. The aforementioned receiver is a correlator receiver. From the discussion on ordinary digital transmission systems, a matched filter instead of the correlator (Section 3.3.5) can be employed to the SS system: the matched filter

takes an impulse response of the time-inverted version of the SS code—a (matched) filter for the baseband signal must be considered separately.

From the previous argument and the fact that the error rate performance (vs.  $E_s/N_0$ ) never depends on pulse waveform (see Section 3.3.5), it is obvious that an SS system shows the same error rate performance as an ordinary transmission system under an assumption of flat noise. If we consider the signal to noise power ratio instead of  $E_s/N_0$  before the despreading process, it takes a low value due to wide bandwidth of the SS signal. The receiver for an FH system consists of frequency dehopping circuits. Error rate performance is the same as the DS system or an ordinary system.

The SS system has characteristics that are different from ordinary systems as follows:

(a) *Low-power spectrum density.* Given a transmit power, the power spectrum density of an SS signal drops by a factor of  $1/N_s$ , where  $N_s$  is the SS factor. This means that an SS signal introduces little interference into other narrow-band systems (however, this assumption is not valid for other wideband systems, where their spectrum bandwidth is comparable to or wider than the SS signal).

The low-power spectral density characteristic of an SS system results in secure communications, since it becomes difficult for a third party to detect the communication by sensing a carrier level with an ordinary narrow-band receiver.

(b) *Immunity against (narrow-band) interference or jamming.* A narrowband (interfering) signal is spectrum spread at the despreading process of an SS receiver (Fig. 6.57). At the output of the receiver (narrow-band) low-pass filter, the interference signal power is reduced by an amount equal to the SS factor. Since SS matched filter is a linear circuit (i.e., a transversal filter whose tap coefficients take chip amplitudes of the SS code), the SS matched filter receiver never spreads the spectrum of the narrow-band interfering signal. However, its effect to suppress the interference is the same as with the correlator receiver: the narrow-band signal level at the output of the matched filter is suppressed owing to the random addition in the filter.

(c) *CDM and CDMA.* If SS codes with low cross correlations are used, SS signals that are transmitted on the same frequency and at the same time can be received with low interference. If orthogonal codes are used, there is no interference between the SS signals (Section 2.1.3.6). As a result, the simultaneous use of radio spectrum on the same frequency and time between different signals, that is, code

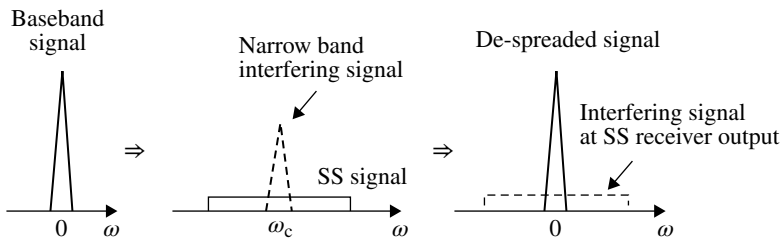
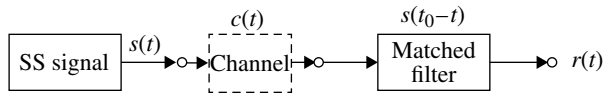


FIGURE 6.57 Spectra in spread-spectrum system.



**FIGURE 6.58** Channel impulse response measurement by matched filtered spread-spectrum system.

division multiplexing (CDM), is possible. If this concept is applied to multiple-access systems, it is called code division multiple access (CDMA). A CDMA system requires no frequency management between users by virtue of using different SS codes as long as the interference between different code signals can be neglected.

(d) *High time resolution.* In general, a narrower pulse shows a higher time resolution. Because of the wide bandwidth of an SS signal, time difference between signals can be measured more accurately than with a narrow-band system as shown in the following. For this purpose, SS codes with a sharp autocorrelation function (e.g., the delta function) are appropriate, as explained in the following. Let us analyze an SS system as shown in Figure 6.58. At the output of the SS signal matched filter, received signal becomes

$$r(t) = s(t) * c(t) * s(t_0 - t) = c(t) * s(t) * s(t_0 - t) \quad (6.21)$$

where  $*$  denotes the convolution integral,  $s(t)$  is the SS signal,  $c(t)$  is the channel impulse response, and  $s(t_0 - t)$  is the impulse response of the matched filter where  $t_0$  is a time constant. Defining a signal  $z(t) \equiv s(t) * s(t_0 - t)$ , it is expressed as

$$z(t) = \int_{-\infty}^{\infty} s(t-x)s(t_0-x)dx \quad (6.22)$$

At a time  $t = t_0 + \tau$ , we have

$$z(t_0 + \tau) = \int_{-\infty}^{\infty} s(y + \tau)s(y)dy \quad (6.23)$$

The right-hand side of Equation 6.23 is the autocorrelation function  $R(\tau)$  of the SS code,  $s(t)$ . Since  $r(t) = c(t) * z(t)$ , if  $s(t)$  shows a sharp autocorrelation, that is,  $|R(\tau)| \ll 1$  for  $\tau \neq 0$  (ideally  $R(\tau) = \delta(\tau)$ , the delta function), then we have the channel impulse response,  $r(t) = c(t)$ . The maximum sharpness of the autocorrelation function is an order of the chip duration. Then an SS system shows a time resolution given by a pulse with the chip duration. The high time resolution characteristic can be useful for ranging distances with a radar. In some radio communication systems, measurement of the channel impulse response becomes important. An SS system can measure it with a high time resolution.

(e) *Rake Receiver.* First, let us consider a conventional system instead of SS systems. A (mobile) radio channel is subjected to multipath fading. In this channel,

a narrow-band signal experiences flat fading, where all frequency components of the signal drop by the same amount at the same time. As a result, a signal level may drop to be below the threshold value for adequate communication. As the signal bandwidth becomes comparable to or wider than the coherence bandwidth of the multipath channel, the signal experiences frequency-selective fading, and the signal level seldom drops below the threshold value (Section 4.4.1), which is an advantage for the wideband signal. However, on the other hand, signal waveform distortion or intersymbol interference due to the frequency-selective fading becomes a problem. In order to solve this problem, automatic channel equalization can be used; however, this requires sophisticated signal processing.

An SS system offers a simple technique effective to mitigate the multipath fading: the technique is called a RAKE receiver (Fig. 6.59) and uses a filter matched to the channel transfer characteristics. The matched filter in the RAKE receiver outputs, at the sampling instant, a signal made by coherently combining the multipath signal components (see Section 3.3.5). Since the multipath signal components are subjected to independent fading, the combined signal has a diversity gain (Section 7.1). Owing to the high time resolution, an SS system yields the channel impulse response necessary for the matched filter (RAKE) receiver. Thus, the SS RAKE receiver achieves the benefit of a wideband transmission with a low effort.

In order to understand the SS system, it is instructive to compare an SS system with a non-SS polar signaling system under the following conditions: same spectrum bandwidth, same data rate, and same average transmit power (Fig. 6.60). Let us denote the data rate and the SS factor as  $f_b$  and  $N_s$ , respectively. To get the same spectrum bandwidth, the pulse duty ratio of the non-SS polar signal must be  $1/N_s$ ; that is, the pulse waveform is the same as the “chip” pulse waveform of the SS system. To have the same average transmit power, the pulse height of the non-SS polar signal must be  $\sqrt{N_s}$  times higher than that of the SS system. Under the aforementioned conditions, the average power spectrum density of both systems is lowered by a factor  $N_s$  when compared with a full-duty polar signaling system. Error rates between the three systems are also the same when a matched filter or correlator receiver is used.

Both wideband systems show the same immunity against a narrow-band signal interference. The low-duty pulse polar signal has an instantaneous power that is higher by a factor of  $N_s$  when compared to the narrow-band signal, assuming the same average power. For the SS system, the narrow-band signal is spectrum spread by a factor  $N_s$ , and the average power of the interfering narrower-band signal is reduced by this amount at the output of a low-pass filter placed at the spectrum despreading circuit.

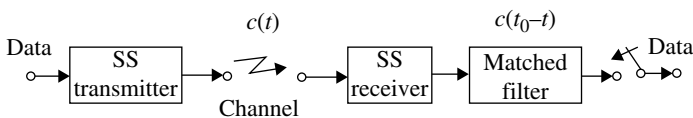
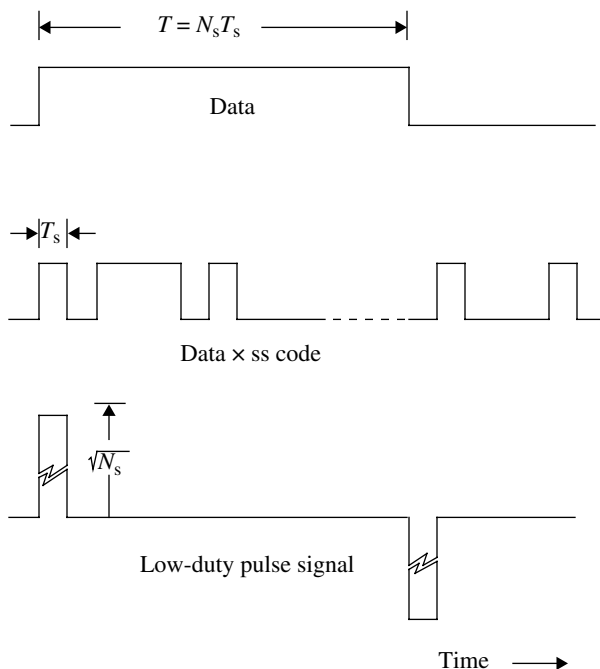


FIGURE 6.59 Rake receiver.



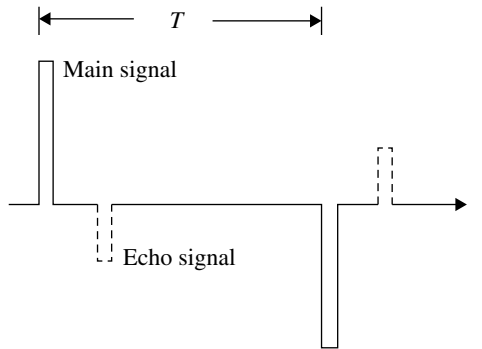
**FIGURE 6.60** SS signal and low-duty polar signal.

Using the low-duty pulse signal, it is seen that  $N_s$  number of signals can be multiplexed in the time domain without any interference (orthogonality). In the SS system, there are  $N_s$  orthogonal codes, for example, the Walsh code (Section 2.1.3), over a spectrum bandwidth expanded by  $N_s$  times. Hence,  $N_s$  signals can also be multiplexed for the SS system. Permitting interference between signals, more signals can be multiplexed using other codes. This is accomplished also with the low-duty pulse system by partially overlapping the pulse in time.

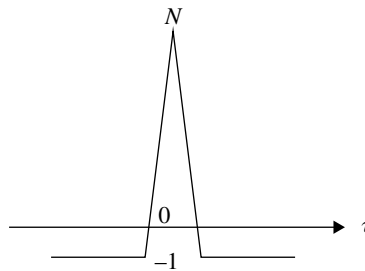
High time resolution with the low-duty pulse system is evident from Figure 6.61. If the autocorrelation function of the SS code is the same as the autocorrelation of the low-duty pulse waveform, then the same results are given for channel impulse response measurement.

As far as we have discussed, we have not seen any differences between the SS system and the low-duty pulse system. However, we have a big difference when peak power is considered. The low-duty pulse system has a peak power  $N_s$  times higher than the SS system. This fact may become important when selecting high-power transmitters.

A slow FH system does not have a high time resolution, even when the average spectrum bandwidth is as wide as a DS system. This result is because the instantaneous or a short-time bandwidth is narrow within the FH system. The FH system has an advantage over the non-FH system: effects of fading are randomized when the frequency hopping range is wide enough compared with the correlation bandwidth of



**FIGURE 6.61** Explanation of channel impulse response for a low-duty pulse system.



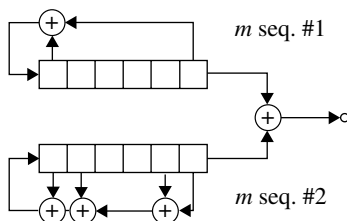
**FIGURE 6.62** Autocorrelation function of the maximum length code.

the fading channel. This is accomplished without intersymbol interference, since the instantaneous bandwidth at the slow FH system is narrow. Thus, the FH technique is an effective countermeasure for shortening the faded periods of a very slow-fading channel. Without this, we must employ a space or frequency diversity technique. A slow FH system is equivalent to a frequency diversity system, where frequency branches are switched periodically according to the FH code.

SS codes are used appropriately in different systems. The maximum length sequence (*m*-sequence) code (Appendix 5.C) is applied to measure the channel, since it shows a sharp autocorrelation function as shown in Figure 6.62. It has side lobes of  $1/N$  time of the peak value, where  $N$  is the length of the sequence. If the measuring distance is long, a long length code is necessary. For a CDM(A) system, codes with low cross correlation are required. The Gold codes have this property; the value of worst cross correlation is kept small. A Gold code generator is shown in Figure 6.63. It is generated by adding specially selected (preferred pair) two *m*-sequences.

Synchronization of SS codes at a correlation receiver is a central issue in the system design. The signal to noise power ratio is quite low when the synchronization is lost, making it difficult to reestablish the SS code synchronization.

Further describing of SS systems is beyond the scope of this book. The reader is referred to [66] and [67] for more details.



**FIGURE 6.63** Gold code generator.

## 6.5 MULTICARRIER TRANSMISSION

In this system, digital signals are transmitted in parallel using multiple carriers with different frequencies. Input signal is serial to parallel converted and used to modulate the multiple carriers. If  $N$  carriers are used, the data rate on a single carrier becomes  $1/N$  times slower compared with a single-carrier system. This serves to mitigate the effect of frequency-selective fading that appears for a case of high data rate transmission. The increase of complexity in modulation and demodulation system is decreased with the fast Fourier transform (FFT) technique (Section 2.4.11). Figure 6.64 shows examples of eye diagrams for multicarrier signals in a heavily dispersive power line channel.

### 6.5.1 Orthogonal Frequency-Division Multiplexing

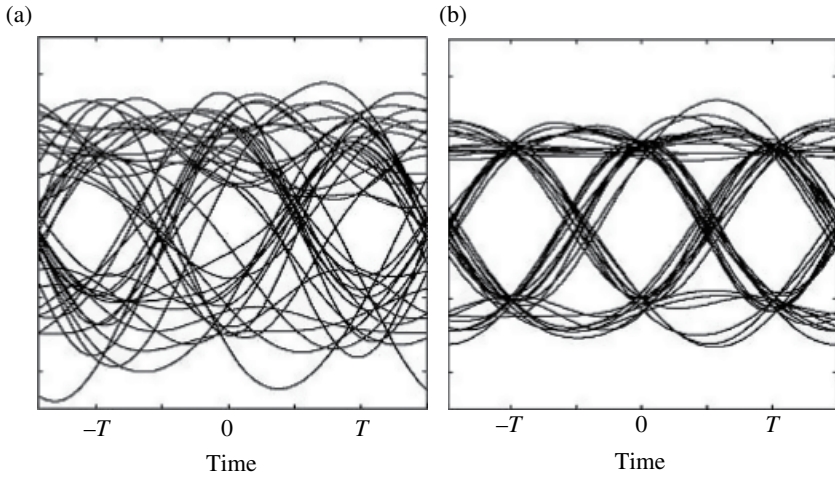
A usual multicarrier system uses subcarrier signals that are orthogonal with each other and hence sow no interference between them [68–70]. This system is called orthogonal frequency-division multiplexing (OFDM) in a wide sense. An OFDM in a wide sense may be spectrum nonoverlapped system or spectrum overlapped system. Orthogonality of a spectrum nonoverlapped OFDM system is easily understood since the subcarrier signals are split by band-pass filters. This is explained mathematically for two subcarrier signals as

$$\int_{-\infty}^{\infty} f_1(t)f_2(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_1(-x)F_2(x)dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_1^*(x)F_2(x)dx = 0$$

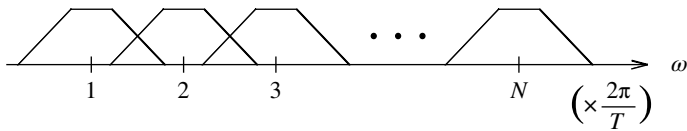
where  $f_1(t) \leftrightarrow F_1(\omega)$ ,  $f_2(t) \leftrightarrow F_2(\omega)$ , and Equations 2.40 and 2.37 are used.

Even with spectra overlap (Fig. 6.65), the components can be split at the demodulator if a proper pulse shaping is made (Appendix 6.A). With overlapping the spectra, the total signal bandwidth can be decreased: an equivalent roll-off factor becomes  $1/N$  times smaller than that for a single modulated signal. The block diagram of the spectrum overlapped multicarrier QAM system [71] is shown in Figure 6.66.

An OFDM system in a narrow sense uses a rectangular pulse signal. Then the condition given with (6.A.6) is satisfied, and the subcarrier signals are orthogonal. We can directly show the orthogonality between two carriers signals from the fact that  $\cos\left(\frac{2\pi}{T}n_1t\right)$ ,  $\sin\left(\frac{2\pi}{T}n_1t\right)$ ,  $\cos\left(\frac{2\pi}{T}n_2t\right)$ ,  $\sin\left(\frac{2\pi}{T}n_2t\right)$  are orthogonal in a time period of  $0 \sim T$ , where  $n_1$  and  $n_2$  are integer.



**FIGURE 6.64** Eye pattern for a channel with high distortion. (a) Number of subcarrier, 49; (b) number of subcarrier, 195.



**FIGURE 6.65** Spectrum of a multicarrier transmission system with spectrum overlap.

For an example if we take

$$f_1(t) = s_1(t) \cos\left(\frac{2\pi}{T} n_1 t\right),$$

$$f_2(t) = s_2(t) \cos\left(\frac{2\pi}{T} n_2 t\right)$$

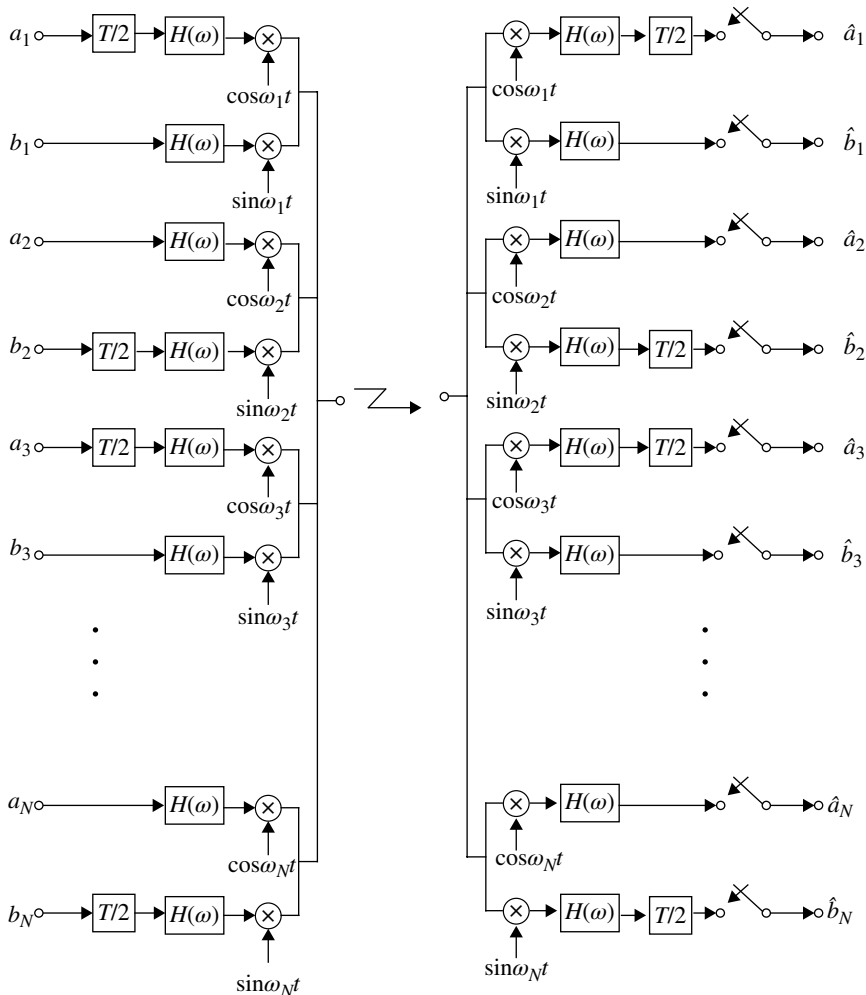
then

$$\int_0^{T_0} f_1(t) f_2(t) dt = \pm \int_0^{T_0} \cos\left(\frac{2\pi}{T} n_1 t\right) \cos\left(\frac{2\pi}{T} n_2 t\right) dt = 0$$

where  $s_i(t)$  ( $i=1, 2$ ) is the rectangular pulse waveform and  $T$  is the symbol time length. Although the spectrum of the rectangular pulse becomes  $T \sin(\omega T/2)/(\omega T/2) e^{-j\omega T/2}$  and spreads infinitely in frequency, the out-of-band spectrum can be neglected when the subcarrier number is sufficiently high.

A benefit of OFDM is that the intersymbol interference can be avoided easily by introducing a guard interval (GI) even in a time-dispersive multipath channel (this GI insertion concept had been proposed for single-carrier transmission systems [72–76]). Actually the copy of the end part of a data block is inserted in the beginning





**FIGURE 6.66** Orthogonal frequency-division multiplexed system.

of the data block (cyclic prefix, CP; Fig. 6.67). If the GI is longer than the time spread of the multipath channel, a preceding symbol never interfere with the succeeding symbols. We explain this qualitatively as follows.

The period of a symbol is integer multiples of those of each subcarrier, and therefore, the subcarrier waveforms are continuous from the CP time period  $-T_g$  to data period  $T$ . The pulse waveforms are also continuous in that time period, since it is rectangular. Thus, the total signal consisted of the CP and data period is continuous in the time period of  $-T_g \sim T$ . Next, we consider a signal that is received at a time delay  $\tau$ , following the main signal. Since the subcarrier signal is a trigonometric waveform, the sum signal of the main and delayed signal is given by

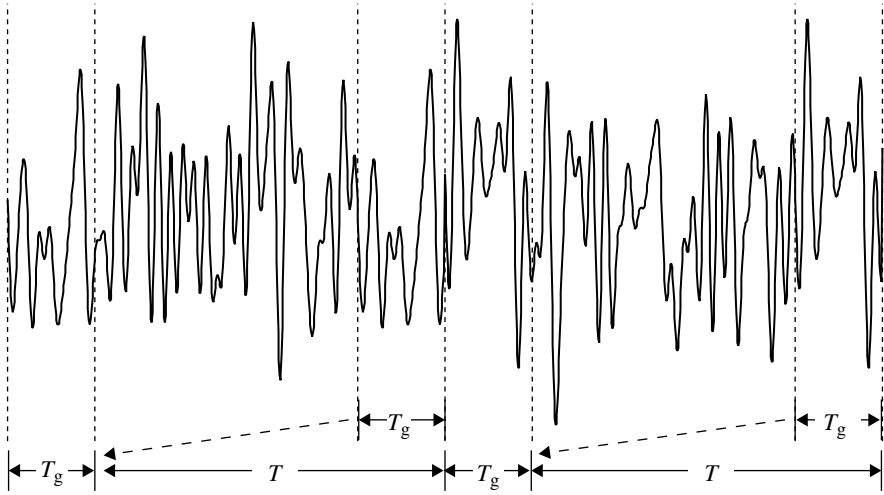


FIGURE 6.67 Guard interval insertion in OFDM signal.

multiplication of a complex constant in a complex baseband expression. Therefore, we have no intersymbol interference: the transmit signal is recovered by compensating the complex constant multiplication.

We will express the previous argument mathematically. Consider  $n$ th representative subcarrier as

$$f_n(t) = a_n p(t) e^{j\omega_n t}, \quad (0 \leq t \leq T)$$

where  $\omega_n = 2\pi/T$ , and  $p(t) = 1$  ( $0 \leq t \leq T$ ) and  $p(t) = 0$  (otherwise).

The signal with the prefix CP is expressed as

$$f'_n(t) = a_n p'(t) e^{j\omega_n t} \quad (-T_g \leq t \leq T)$$

where  $p'(t) = 1$  ( $-T_g \leq t \leq T$ ). We express the impulse response of the channel as

$$h(t) = \sum_{m=0}^M h_m \delta(t - \tau_m)$$

Then the received signal is given as

$$\begin{aligned} r_n(t) &= \sum_{m=0}^M a_n h_m p'(t - \tau_m) e^{j\omega_n(t - \tau_m)} \\ &= a_n e^{j\omega_n t} \sum_{m=0}^M h_m p'(t - \tau_m) e^{-j\omega_n \tau_m} \quad (-T_g \leq t \leq T + \tau_m) \end{aligned}$$

where  $\tau_m < T_g$  and  $p'(t - \tau_m) = 1$  for  $0 \leq t \leq T$ . Thus we get

$$r_n(t) = a_n p(t) e^{j\omega_n t} H(\omega_n) \quad (0 \leq t \leq T)$$

where the transfer function  $H(\omega_n)$  of the channel is given as

$$H(\omega_n) = \sum_{m=0}^M h_m e^{-j\omega_n \tau_m}$$

The constant for channel compensation becomes  $G(\omega_n) = H^*(\omega_n) / |H(\omega_n)|^2$ . The channel transfer function  $H(\omega)$  can be estimated using the discrete Fourier transform (DFT) result of the received reference signal and that of the reference signal.

The insertion of CP signal is actually carried out after the total multicarrier signal is generated with the inverse FFT.

The benefit of GI insertion for a multicarrier signal transmission appears also in a digital broadcasting system: a so-called single-frequency network (SFN) can be realized [77], where multiple broadcasting stations send the same signal at the same time to cover wide service areas for efficient use of the radio frequencies. If the time arrival differences from different broadcasting stations are within the GI,  $T_g$ , the interference between the received signals are avoided. Such a technique is not applicable to a conventional analog broadcasting system.

Channel efficiency reduces by a factor of  $T/(T + T_g)$  due to the GI insertion. The required length of  $T_g$  depends on multipath delays of the channel. For a given  $T_g$ , a longer symbol duration  $T$  is required to increase the channel efficiency. For a longer  $T$ , the number of subcarriers should be increased to keep the same transmission speed. This results in the increase of FFT points and hence memory size to process multicarrier signals.

To avoid just intersymbol interference, we need not send the CP signal but just make a signal blank during GI. The benefit of the CP insertion appears at the receiver. First, we can enjoy a timing margin to start sampling of the received signal: since the CP is attached such that the transmit signal is cyclically repeated, any sampling starting timing is allowed, as long as the timing is within the CP signal time period. The timing correction is made automatically when we equalize the channel distortion with FFT in the frequency domain. If the channel is free from distortion and the timing error is  $\Delta t$ , then the estimated channel transfer function becomes  $H(\omega) = e^{-j\omega \Delta t}$ . This time delay is corrected by channel transfer compensation.

The second benefit with CP is to serve for establishing the symbol timing at a receiver. The multicarrier signal as a whole becomes like a noise and have a sharp autocorrelation function. By taking correlation between a received signal and its delayed version with a time delay of  $T$ , we get a sharp peak at the CP signal time period and know the starting time point of the symbol.

The other benefit with CP appears when establishing the frequency synchronization at a receiver. Let us assume the transmit signal as  $b(t)e^{j\omega_c t}$  and the receiver local oscillator frequency as  $\omega_c - \Delta\omega$ . Then the received baseband signal is given as

$r(t) = b(t)e^{j\Delta\omega t}$ . We detect  $\Delta\omega$  to compensate for the local oscillator frequency error. Delaying  $r(t)$  by time  $T$  and taking its complex conjugate and multiplying it to  $r(t)$ , we have

$$r(t)r^*(t-T) = b(t)e^{j\Delta\omega t} \cdot b^*(t-T)e^{-j\Delta\omega(t-T)} = |b(t)|^2 e^{j\Delta\omega T}$$

$$(nT < t < nT + T_g)$$

Detecting the phase of this signal, we get  $\Delta\omega T = \frac{\Delta\omega}{f_0}(f_0 = 1/T)$ . To decrease the noise effect on the result, we take a time average during time period of  $T_g$ .

Pilot signals are used for purposes such as channel characteristic estimation in OFDM systems. The channel characteristics change in time as well as in frequency. The pilot signals therefore should be placed to cope with the change. To this end, the sampling theorem both in the time and frequency domains (Section 2.4.1) should be satisfied. The pilot signal decreases channel efficiency, and it should be designed compromising the channel efficiency and the adaptation speed. Figure 6.68 shows an example of a pilot signal design. Pilot signals are placed discretely in the time domain and frequency domain. Channel characteristics in a subcarrier at every time are estimated through interpolation with the pilot signals on the subcarrier. Then the interpolation in the frequency domain is carried out for each subcarrier.

To get an insight into multicarrier transmission, we compare multicarrier systems with single-carrier systems in the following. The multicarrier system mitigates intersymbol interference owing to slow parallel transmission and suppresses it with GI insertion. In contrast to this, single-carrier system should require equalization of channel distortion to combat heavy intersymbol interference. The complexity of the equalization increases with the increase of transmission speed, since the intersymbol interference spans over a higher number symbol period. If we use a frequency-domain equalization method with FFT (Section 7.3.7), the signal processing complexities are comparably equal between both systems. A disadvantage for a multicarrier system against a single carrier system is its high peak to average power ratio (PAPR): the PAPR value increases with the number ( $N$ ) of multiplexed signals, since the average power of independent signal of equal average power ( $P_0$ ) becomes  $NP_0$  and the peak

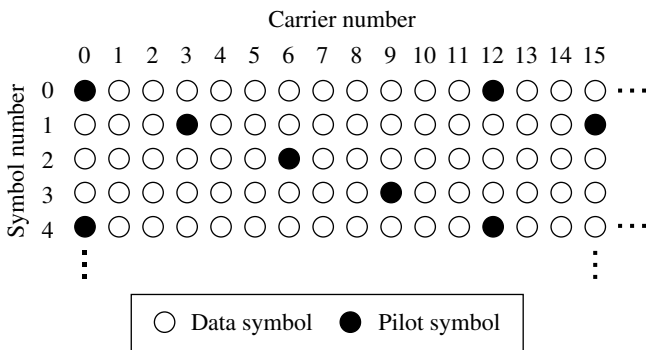
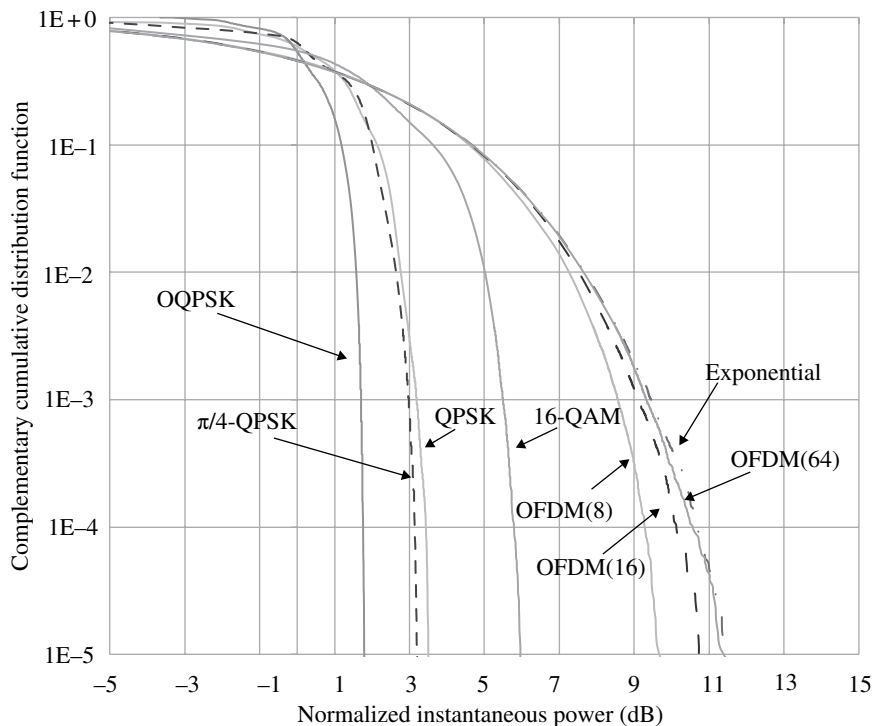


FIGURE 6.68 Insertion of pilot signal in OFDM system.



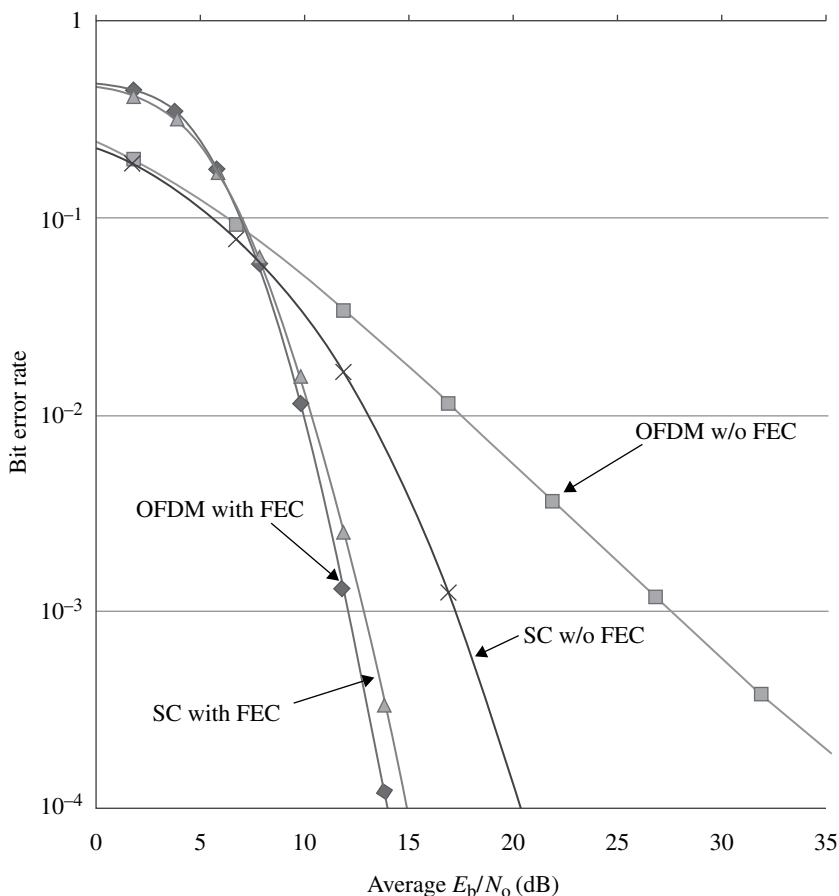
**FIGURE 6.69** Complementary cumulative distribution function of normalized power.

power increases as  $(N\sqrt{P_0})^2 = N^2P_0$ . The probability of the peak power appearance, however, decreases with the increase of  $N$ . Probabilities of instantaneous signal power are sometimes shown with a CCDF. Figure 6.69 shows the CCDF examples for OFDM signals (narrow sense) and other modulation methods. The horizontal axis shows instantaneous power normalized with its average power, and the vertical axis corresponds to probabilities that the normalized instantaneous power exceeds the horizontal value. We can see that the OFDM signal instantaneous power nearly obeys an exponential distribution (Eq. 2.80), which is given for a two-dimensional Gaussian noise, when the multiplexing number is increased to 64. A high PAPR value of a signal makes it difficult for the signal to achieve a low nonlinear distortion and a high power efficiency at a power amplifier. To reduce the PAPR, many techniques have been developed (Section 8.6.3).

Next, we discuss error rate performance. Both two systems give the same error rate in a static condition if the two systems use the same modulation method. Even in a fading condition, the same result holds, as long as the fading channel is free from distortion, that is, flat fading. In a frequency-selective fading channel, the single-carrier system suffers from heavy intersymbol interference, and its error rate performance deteriorates. Contrary to this, the multicarrier system shows the same error rates as for a flat fading, as long as the number of multiplexed signal is sufficiently high. If we introduce to the single-carrier system a channel distortion

equalizer based on the minimum mean square criterion, the error rate performance becomes much better than that of the multicarrier system as shown in Figure 6.70, owing to the frequency diversity effect of the equalizer. However, the difference in the error rate performance almost disappears when an error correction is introduced as shown in the same figure. This is because the original error rates are almost the same between the two systems. The error rate performance reversal between the two methods before and after the introduction of the error correction means that the error correction capability favors the multicarrier system: probably the errors become more burst for the single-carrier system than for the multicarrier system.

An example of a multicarrier transmission system is a data communication system at HF or short-wavelength band [78] for mobile communications. At HF band, the frequency-selective fading becomes severe since the delay difference between multipath signals becomes large due to an ionospheric propagation circumstance. Channel distortion caused by the frequency-selective fading is avoided with the multicarrier



**FIGURE 6.70** Bit error rates comparison between multicarrier and single-carrier systems.

transmission systems. Multicarrier systems are highlighted in recent years and introduced to wireless local area networks (Section 9.8) and cellular systems (Section 9.7.6). Reference [79] discussed for the first a multicarrier system for application to land mobile radio communications. The history of research and development of the multicarrier systems is described in [80]. Performance of multicarrier systems is well investigated in [81].

### 6.5.2 Generation of Multicarrier Digital Signal

Digital signal processing such as filtering, equalization, and modulation is sometimes carried out efficiently in the frequency domain, where signal is transformed into the frequency domain with discrete Fourier transform (DFT, Section 2.4.10). DFT is carried out efficiently with the fast Fourier transform.

DFT is the Fourier transform of a time-sampled periodic signal. DFT is understood as the Fourier power series expansion with orthogonal trigonometric functions in a given time period  $T$  (Section 2.1.2). The frequency components become line spectra at frequencies of integer multiples of  $\Omega = 2\pi/T$ . Since the signals with these frequencies are orthogonal with each other in time period of  $0 \leq t \leq T$ , we call subcarriers with these frequencies  $\omega_n$  as orthogonal subcarriers.

A digital multicarrier signal with  $N$  subcarriers is expressed as,

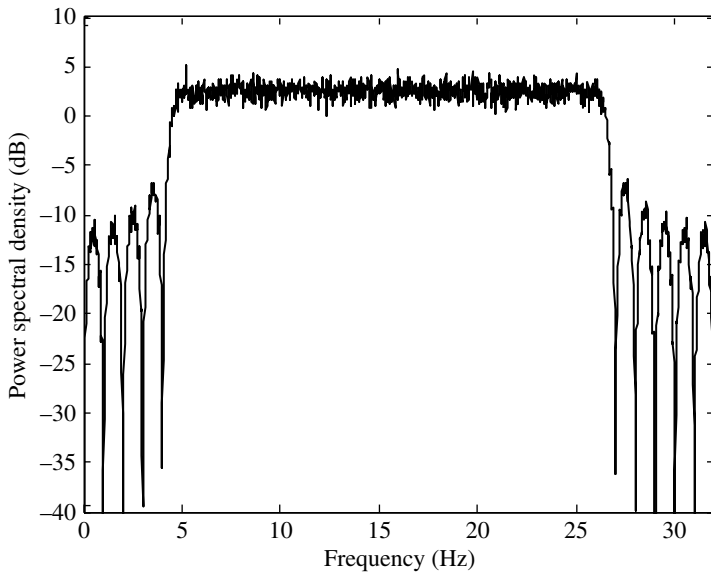
$$f(t) = \sum_{i=0}^{N-1} \sum_{k=0} a_{ki} p(t - iT) e^{j\omega_k t} \quad (6.24)$$

where  $a_{ki}$  denotes the digital (complex) signal conveyed with  $k$ th subcarrier,  $p(t)$  the pulse waveform,  $\omega_k$  the frequency of  $k$ th subcarrier, and  $T$  the symbol period. We sample this signal with sampling period  $T_s$ . If the sampling period  $T_s$  is sufficiently shorter than symbol period, the sampling theorem is satisfied for sampling of the pulse waveform  $p(t)$ . However, the sampling of subcarriers amplitude modulated with  $p(t)$  may cause aliasing error since the modulated subcarrier has frequency components higher than  $\omega_s/2$  (where  $\omega_s = 2\pi/T_s$ ) required by the sampling theorem. In this case, we do not use such a subcarrier by making  $a_{ki} = 0$ . For an example of such a case, Figure 6.71 shows spectrum of a multicarrier signal with a rectangular pulse waveform. Due to the rectangular pulse waveform, the spectrum spreads into out of band frequency region, although its intensity decays as the frequency departure from center frequency. The spectrum spread decreases relatively with increase of number of subcarriers. Here we assume that the sampling theorem is fulfilled even when all the subcarriers are used.

We truncate the pulse waveform within  $L$  symbol period ( $0 \leq t \leq LT$ ). That is,  $p(t) = 0$  ( $t < 0, t > LT$ ).  $L$  is decided considering the error in frequency spectrum. We can introduce a window function for the time truncation (Section 2.4.9). The sampled signal is written as

$$f_s(t) = \sum_{n=0}^{N-1} f(nT_s) \delta(t - nT_s)$$

where  $N$  is a signal block size.



**FIGURE 6.71** Spectrum of multi-carrier system with rectangular pulse ( $N_{\text{sub}} = 22$ ,  $N = 32$ ,  $f_s = N/s = 32$  Hz).

A straight forward method to generate a multicarrier signal by digital signal processing is to apply the digital signal into a pulse waveform generating digital filter, multiply the output signal to a corresponding subcarrier (modulation), and add the modulated subcarrier signal. If all the digital signal processing is done at sampling period  $T_s$ , the total number of complex multiplying during a symbol period  $T$  becomes  $LN^2 + N^2$ , where  $N = T/T_s$  is the number of samplings per symbol period. The method can be applied to any multicarrier system with an arbitrary subcarrier frequency and pulse waveform.

If the pulse waveforms are the same for all the subcarrier and the subcarrier frequencies are an integer multiple of the symbol frequency, that is,  $\omega_k = k\Omega$  ( $\Omega = 2\pi/T$ ), the number of multiplication can be reduced as follows:

The signal given with Equation 6.24 is expressed at  $i$ th symbol time as

$$\begin{aligned} f(t) &= \sum_{n=0}^{L-1} \sum_{k=0}^{N-1} a_{k(i-n)} p(t - iT + nT) e^{j\omega_k t} \\ &= \sum_{n=0}^{L-1} p(t - iT + nT) \sum_{k=0}^{N-1} a_{k(i-n)} e^{j\omega_k t} \quad (iT \leq t \leq (i+1)T) \end{aligned} \quad (6.25)$$

We use as possible time-continuous expression of the signal for a better intuitive understanding. Letting  $t = t' + iT$ , we rewrite the above equation as

$$f(t) = \sum_{n=0}^{L-1} p(t' + nT) \sum_{k=0}^{N-1} a_{k(i-n)} e^{j\omega_k t'} \quad (0 \leq t' \leq T)$$



where  $e^{j\omega_k iT} = e^{j\omega_k T} = 1$  is used. (This relation will appear often later. We do not point out this as it appears.) If we let

$$p_n(t') = p(t' + nT), \quad A_{i-n}(t') = \sum_{k=0}^{N-1} a_{k(i-n)} e^{j\omega_k t'}$$

we have

$$f(t') = \sum_{n=0}^{L-1} p_n(t') A_{i-n}(t')$$

This equation means that  $f(t')$  is given by a convolution of  $p_n(t')$  and  $A_{i-n}(t')$ , that is, by applying signal  $A_{i-n}(t')$  to a digital filter with coefficients  $p_n(t')$  at a sampling rate of  $1/T$ . The impulse response of the digital filter becomes

$$p_m(nT) = p(mT_s + nT) \quad (m = 0, 1, 2, \dots, N-1; \quad n = 0, 1, 2, \dots, L-1)$$

The term  $A_{i-n}(t') = \sum_{k=0}^{N-1} a_{k(i-n)} e^{j\omega_k t'}$  expresses the inverse Fourier transform of a signal having discrete frequency components  $\omega_k$  and levels  $a_{k(i-n)}$ . If we use the DFT expression, we have

$$A_{i-n}(l) = \sum_{k=0}^{N-1} a_{k(i-n)} e^{j2\pi kl/N} \quad (l = 0, 1, \dots, N-1)$$

Here the sampling period is the symbol duration  $T = NT_s$ . Thus, we can generate a multicarrier signal by doing an IDFT of digital data signals, applying the output time-discrete signals to digital filters, whose coefficients takes the sampled values of the pulse waveform, and multiplexing the filtered signals in the time domain (parallel to serial conversion). The signal generator with this method is shown in Figure 6.72.

The number of multiplication per symbol period becomes  $LN + N^2$  ( $LN$  for filtering,  $N^2$  for DFT). If we use the FFT, the number reduces to  $LN + M \log_2 N$ , which is smaller than  $LN^2 + N^2$  with the straightforward method. When the pulse waveform is rectangular (OFDM in a narrow sense), the digital filter can be omitted, since  $p_m(nT) = 0$  ( $n \neq 0$ ) and  $p_m(0) = 1$ .

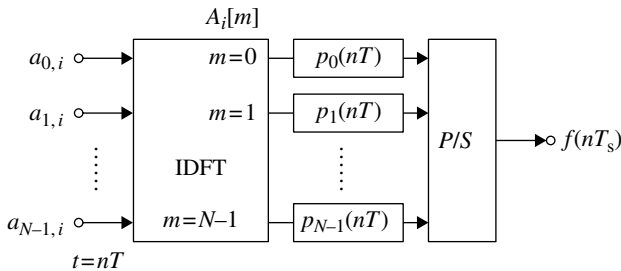


FIGURE 6.72 Multicarrier signal generation with IDFT.

The signal given by Equation 6.24 can be generated in a different way by applying the digital signals to band-pass filters as

$$f(t) = \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} a_{ki} \delta(t - iT) * h_k(t)$$

where  $h_k(t) = p(t)e^{j\omega_k t}$  is the impulse response of the band-pass filter of which the center frequency is  $\omega_k$ . The same transmitter can be obtained from a frequency-domain expression of the above equation [82]. The key point of the argument there is decomposition of  $z$ -transform of  $p(t)$ ,  $P(z) = \sum_{n=0}^{LN-1} p_n z^{-n}$  ( $z = e^{j\omega T_s}$ ), as

$$P(z) = \sum_{m=0}^{N-1} z^{-n} \sum_{i=0}^{L-1} p_{m+iN} z^{-iN} \quad (p_{m+iN} = p(mT_s + iNT_s))$$

This treatment is equivalent to modification of the expression from  $p(t)$  ( $0 \leq t \leq LT$ ) to  $\sum_{n=0}^{L-1} p(t' + nT)$  ( $0 \leq t' \leq T$ ) in the above derivation.

We used so far a complex expression of signals. In an actual implementation, we should return to use real number expression. If we take the real part of signals with the same sampling frequency, the sampling theorem is not fulfilled. We should take the real part after upsampling (by putting zeros between two samples: Section 2.4.7) followed by low-pass filtering to cut the higher harmonics. Then we upconvert the signal to a desired frequency  $\omega_c$ . If we use amplitude modulation for the upconversion, we have a double-sideband (DSB) spectrum at  $\omega_c$ , and therefore, we should use single-sideband modulation. For SSB modulation, we need the Hilbert transform of the baseband signal  $s(t)$  and generate the SSB signal as (Section 2.3.4).

$$f(t) = s(t) \cos \omega_c t \pm \hat{s}(t) \sin \omega_c t$$

Since each subcarrier is already modulated and written as

$$\begin{aligned} s(t) &= \sum_{k=0}^{N-1} [x_k(t) \cos \omega_c t - y_k(t) \sin \omega_c t] \\ &= \text{Re} \left\{ \sum_{k=0}^{N-1} [x_k(t) + jy_k(t)] e^{j\omega_k t} \right\} \end{aligned}$$

the Hilbert transform  $\hat{s}(t)$  of  $s(t)$  is given as (Section 2.3.4, Example 2.10)

$$\begin{aligned} \hat{s}(t) &= \sum_{k=0}^{N-1} [x_k(t) \sin \omega_c t + y_k(t) \cos \omega_c t] \\ &= \text{Im} \left\{ \sum_{k=0}^{N-1} [x_k(t) + jy_k(t)] e^{j\omega_k t} \right\} \end{aligned}$$

### 6.5.3 Demodulation of Multicarrier Signals

Noise and interference between signals cause received signal quality degradation in a communication system. In a multicarrier system, interference between different subcarrier signals may occur as well as intersymbols in a subcarrier. These matters are put away in the argument so far.

With respect to noise and the inter-symbol interference, a matched filter receiver that meets the Nyquist I condition together with a transmit filter becomes optimum (Section 3.3.6). At the same time, the intersubcarrier interference should be suppressed. The conditions for this are described in Appendix 6.A and are assumed in the following. We never consider the effect of noise, since the effect of noise is the same as in an ordinary digital transmission system due to assumption of the matched filter receiver.

Received signal is expressed with Equation 6.25. A straight forward method to demodulate the signal is to do quadrature demodulation of each subcarrier signal and pass the output to a low-pass matched filter. This method is commonly used when the receiver is implemented with analog circuits. Here, we show the other method that is suited to digital signal processing, and the complexity is reduced. The principle of demodulation is described first, and then the circuit implementation is given. We consider a matched filter receiver. A correlator receiver as the other method gives the same result at the signal decision output (Section 3.3.5).

The impulse response of a matched filter for a subcarrier with a carrier frequency of  $\omega_m$  becomes

$$h_m(t) = p(LT - t)e^{j\omega_m t} \quad (0 \leq t \leq LT)$$

Applying the received signal  $y(t)$  to the matched filter, we have

$$\begin{aligned} r_m(t) &= \int_{-\infty}^{\infty} y(x)h_m(t-x)dx \\ &= \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} a_{ki} e^{j\omega_m t} \int_{-\infty}^{\infty} p(x-iT) e^{j\omega_k x} p(LT+x-t) e^{-j\omega_m x} dx \end{aligned}$$

At time instant  $t = nT$ , it becomes

$$r_m(nT) = \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} a_{ki} e^{j\omega_m nT} \int_{-\infty}^{\infty} p(x-iT) e^{j\omega_k x} p(LT+x-nT) e^{-j\omega_m x} dx$$

and  $r_m(nT)$  becomes zero except for  $k = m$  and  $i = n - L$  owing to orthogonality in the time and frequency domain. Therefore, we have

$$r_m(nT) = a_{m(n-L)} E_b \quad (m = 0, 1, 2, \dots, N-1)$$

where  $E_b = \int_{-\infty}^{\infty} p^2(t)dt$  is the pulse energy. The transmit digital signal is obtained as  $a_{m(n-L)} = r_m(nT)/E_b$ . Actually, the decision making is done since noise exists.

We will derive the receiver implementation in the following. The output of the matched filter becomes

$$\begin{aligned} r_m(t) &= \int_{-\infty}^{\infty} y(x)p(LT+x-t)e^{j\omega_m(t-x)}dx \\ &= e^{j\omega_m t} \int_{-\infty}^{\infty} y(x)p(LT+x-t)e^{-j\omega_m x}dx \end{aligned}$$

At time instants  $t = nT$  ( $n \geq L$ ), we have

$$r_m(nT) = \int_{-\infty}^{\infty} y(x)p(LT+x-nT)e^{-j\omega_m x}dx$$

Since  $p(t) = 0$  ( $t < 0, t > LT$ ), it becomes

$$r_m(nT) = \int_{(n-L)T}^{nT} y(x)p(x-(nT-LT))e^{-j\omega_m x}dx$$

Letting  $t = x - kT$  ( $k = n - L$ ), we have

$$r_m(nT) = \int_0^{LT} y(t+kT)p(t)e^{-j\omega_m t}dt$$

We divide the integration into  $L$ -region of period  $T$ , and let  $t = t' + iT$  ( $0 \leq t' \leq T$ ), and then we get

$$\begin{aligned} r_m(nT) &= \sum_{i=0}^{L-1} \int_0^T y(t'+iT+kT)p(t'+iT)e^{-j\omega_m t'}dt' \\ &= \sum_{i=0}^{L-1} \int_0^T y(t'+iT-LT+nT)p(t'+iT)e^{-j\omega_m t'}dt' \end{aligned}$$

Letting  $t = lT_s$ , we have the sampled signal as

$$r_m(nT) = \sum_{l=0}^{N-1} d_n[l]e^{-j2\pi n l/N} \quad (6.26)$$

where

$$d_n[l] = \sum_{i=0}^{L-1} y(lT_s + iT - LT + nT)p(lT_s + iT) \quad (6.27)$$

Equation 6.27 shows a signal of a time length  $T$ , which is produced by the  $L$ -times addition of the  $L$ -division of the received signal of a length  $LT$  multiplied with the

pulse waveform  $p(t)$ . The  $L$ -times weighted addition is implemented as a digital filter with a clock period  $T$ .

The structure of the filter is the same as the transmit filter if the pulse waveform  $p(t)$  is symmetrical with respect to  $t = L/2$ . Equation 6.26 shows a DFT. The receiver function is summarized as follows: the received signal is sampled at a period  $T_s$ , then serial to parallel transformed with a block of  $N$  samples, filtered on each parallel signal at period  $T$ , and subjected to DFT, and a decision is made for the  $N$  DFT output signals. This operation of the receiver is the time inverse of that of the transmitter. The number of multiplications is therefore the same for the transmitter and the receiver.

The Fourier transform of a signal with time period  $T$ , which is made by the  $L$ -times addition of the  $L$ -divided received signal, instead of the Fourier transform of the  $LT$  time length received signal, becomes possible, since we are interested in the special frequencies  $\omega_k = k\Omega$  ( $\Omega = 2\pi/T$ ). Thus, the computation complexity is reduced by decreasing the block size of DFT from  $LN$  (the pulse waveform length) to  $N$  (symbol duration) (Section 2.4.10).

The same pulse waveform and the same symbol timing of subcarriers are assumed in the argument so far. A multicarrier system with offset QAM uses subcarrier whose inphase and quadrature-phase components are time shifted by one-half of a symbol duration. In such a system, it is effective that the IDFT is carried out twice the symbol duration and the filter is operated in the time domain at two times higher rates of the symbol frequency [71].

The transmitter and receiver with a digital signal processing as discussed earlier can be applied to an analog signal multicarrier system [82]: the sampled analog signal is considered as digital signal, and the pulse waveform as a band-limiting filter for the analog signal, as long as the band-limiting filter never distorts the analog signal.

We assumed that subcarrier frequencies take an integer multiple of the symbol frequency. Contrary to this, subcarrier frequencies are shifted from the integer multiple of the symbol frequency as in a multicarrier system for a public telephone line transmission. In those systems, the signal processing becomes a little complex with the subcarrier frequency shift.

## 6.6 SINGLE-CARRIER FREQUENCY-DIVISION MODULATION

A conventional modulated signal has a spectrum lumped at the carrier frequency. The single-carrier frequency-division modulated signal discussed recently shows a spectrum that is divided into several separated frequency regions and spreads by putting null spaces in the frequency domain. This system has benefits as follows: (i) a frequency-division multiplexing (or access) becomes possible by inserting another signal into the null spaces and (ii) frequency diversity effect is obtained due to the spectrum spread as in a spectrum spread system in a frequency-selective fading channel.

In the following, generation of the single-carrier frequency-division modulated is discussed. First, we describe a generation method of a conventional single-carrier digital signal with digital signal processing. The equivalent base band expression of a single-carrier signal is expressed as

$$f(t) = \sum_{i=0}^{M-1} a_i p(t - iT)$$

where  $a_i$  is a complex digital signal,  $p(t)$  the pulse waveform,  $T$  the symbol duration, and  $M$  the total number of transmit data symbols. We truncate  $p(t)$  in an  $L$  symbol time duration ( $0 \leq t \leq LT$ )  $a_i = 0$  for  $i \geq M - L$  is understood; the signal stays in a time period  $0 \leq t \leq MT$ . Taking sampling of  $f(t)$  at a period of  $T_s (= T/N)$ , we have  $f(nT_s)$  ( $n=0, 1, 2, \dots, MN-1$ ). Doing DFT of this signal, we get

$$\begin{aligned} F(m\Omega) &= \sum_{n=0}^{MN-1} \sum_{i=0}^{M-1} a_i p(nT_s - iT) e^{-j2\pi mn/MN} \\ &= \sum_{n=0}^{MN-1} \sum_{i=0}^{M-1} a_i p(nT_s) e^{-j2\pi mn/MN} e^{-j2\pi mi/M} \\ &= \sum_{n=0}^{LN-1} p(nT_s) e^{-j2\pi mn/MN} \sum_{i=0}^{M-1} a_i e^{-j2\pi mi/M} \\ &= P(m\Omega) A(m\Omega) \quad (m=0, 1, 2, \dots, MN-1) \end{aligned}$$

where

$$\Omega = 2\pi/MT, P(m\Omega) = \sum_{n=0}^{MN-1} p(nT_s) e^{-j2\pi mn/MN} = \sum_{n=0}^{LN-1} p(nT_s) e^{-j2\pi mn/MN}, A(m\Omega) = \sum_{i=0}^{M-1} a_i e^{-j2\pi mi/M}$$

$P(m\Omega)$  is given by DFT of  $p(t)$  at a sampling period  $T_s$ , and  $A(m\Omega)$  is given by DFT of  $a_i$  at a sampling period  $T$ .

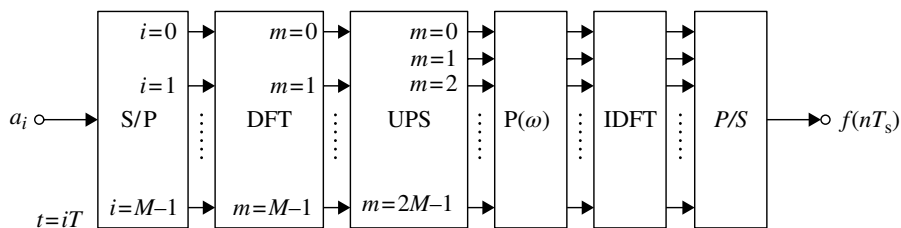
Letting  $m = m' + kM$  ( $m' = 0, 1, 2, \dots, M-1$ ;  $k = 0, 1, 2, \dots, N-1$ ), we have

$$A(m'\Omega + kM\Omega) = \sum_{i=0}^{M-1} a_i e^{-j2\pi m'i/M}.$$

Thus, instead of the calculation for  $m = 0, 1, 2, \dots, MN-1$ , the calculation for  $m' = 0, 1, 2, \dots, M-1$  ( $k=0$ ) is sufficient: the other calculation ( $k \neq 0$ ) is made by copying the calculation results for the same  $m'$ . This fact corresponds to a  $N$  times upsampling in the time domain (Section 2.4.9).

Taking the IDFT of  $F(m\Omega)$ , we have the time-domain signal

$$f(nT_s) = \frac{1}{MN} \sum_{m=0}^{MN-1} F(m\Omega) e^{j2\pi mn/MN} \quad (n = 0, 1, 2, \dots, MN-1)$$



**FIGURE 6.73** Single carrier signal generation with DFT/IDFT.

A signal generation circuit with DFT/IDFT for  $N=2$  is shown in Figure 6.73.

The total number of complex multiplication with a straightforward calculation in the time domain becomes  $(M-L)LN$  ( $LN$  per symbol multiplied with  $M-L$  symbols). Using DFT, this number becomes  $(MN)^2 + M^2 + MN$  ( $M^2$  for  $A(m\Omega)$ ,  $MN$  for  $P(m\Omega)A(m\Omega)$ , and  $(MN)^2$  for IDFT). If we use FFT, we have  $MN \log_2(MN) + M \log_2 M + MN$ . With increase in the pulse truncation Length  $L$ , the effect of FFT on relative reduction of the total number of complex multiplexing is enhanced.

The frequency-division single carrier signal can be generated as follows.

Input digital signals are converted into frequency region with DFT, the frequency domain signal is multiplied with the pulse spectrum, each frequency component is moved to the desired frequencies leaving vacant components at the original component frequencies, and the frequency domain signal is transformed into a time-domain signal with IDFT. In this process, if we put the vacant space periodically between the neighboring frequency components, the small PAPR, a benefit of a single-carrier system is kept the same as the original single-carrier signal. This is because  $N$ -zero insertion between the frequency component samples results in repeating the original signal in the time domain  $N$  times as explained in Section 2.4.7. In this case, we need not make DFT with  $N$  times higher number block size: a signal is repeated  $N$  times in the time domain.

We can multiplex other signal into any vacant frequency spaces. When we recover the original signal at a receiver, after making DFT of the received signal and moving the frequency components into the original positions, then we transform them into the time domain with IDFT. For a signal with periodic zero insertion in the frequency domain, we need not make DFT and IDFT, but just combine the  $N$  times repeated signal in the time domain.

The frequency-division single-carrier system is discussed for use in the uplink of the long-term evolution (LTE; Section 9.7.6) system [83]. This system however seems not to be adopted in commercial systems. The reason for this is probably that dynamic scheduling system, which assign frequency channels adaptively for users with a good channel quality (multiuser diversity, Section 9.1.3), gives a higher system throughput than the frequency-division single-carrier system, which improves average throughput for a user with frequency diversity effect given by the spectrum spread.

## APPENDIX 6.A MATHEMATICAL PRINCIPLE OF ORTHOGONAL FREQUENCY-DIVISION MULTIPLEXING

Consider two modulated signals expressed in a complex form:

$$z_1(t) = [a_1 x_1(t) + j b_1 y_1(t)] e^{j \Delta \omega_1 t},$$

$$z_2(t) = [a_2 x_2(t) + j b_2 y_2(t)] e^{j \Delta \omega_2 t},$$

where  $a_i$  and  $b_i$  are information-bearing signals,  $x_i(t)$  and  $y_i(t)$  are waveforms, and  $\omega_i$  are carrier frequencies for  $i = 1, 2$ . All of these are real valued. Actual signals are given by the real part of  $z_i(t)$ . The two signals are summed and transmitted through a channel. We will find (orthogonal) conditions so that the two signals can be separately received. Demodulation of the signal includes frequency translation into a baseband signal and filtering. Letting  $z(t) = z_1(t) + z_2(t)$ , the terms including  $a_i$  and  $b_i$  are given by

$$\operatorname{Re} \left[ \int_{-\infty}^{\infty} z(t) e^{-j \omega_i t} x_i(t) dt \right], \text{ and } \operatorname{Im} \left[ \int_{-\infty}^{\infty} z(t) e^{-j \omega_i t} y_i(t) dt \right]$$

respectively, where we assume a correlation receiver instead of filtering (correlation and matched filtering are equivalent; see Section 3.3.5).

The conditions for any  $a_i$  and  $b_i$  to be separated at a receiver without interchannel interference are

$$\begin{aligned} \int_{-\infty}^{\infty} x_1(t) x_2(t) \cos \Delta \omega t dt &= 0, \\ \int_{-\infty}^{\infty} y_1(t) y_2(t) \cos \Delta \omega t dt &= 0, \\ \int_{-\infty}^{\infty} x_1(t) y_2(t) \sin \Delta \omega t dt &= 0, \\ \int_{-\infty}^{\infty} y_1(t) x_2(t) \sin \Delta \omega t dt &= 0 \end{aligned} \tag{6.A.1}$$

where  $\Delta \omega = \omega_2 - \omega_1$ .

From Equation 6.A.1, we have

$$\begin{aligned} &X_1(\Delta \omega) * X_2(\Delta \omega) + X_1(-\Delta \omega) * X_2(-\Delta \omega) \\ &= X_1(\Delta \omega) * X_2(\Delta \omega) + X_1 * (\Delta \omega) * X_2 * (\Delta \omega) \\ &= 2 \operatorname{Re} [X_1(\Delta \omega) * X_2(\Delta \omega)] = 0, \end{aligned} \tag{6.A.2}$$

where  $x_i(t) \leftrightarrow X_i(\omega)$ ,  $*$  between variables denotes the convolutional integral, and  $X(-\omega) = X^*(\omega)$  for real  $x(t)$  is used.



Similarly, we have

$$\begin{aligned}
 & Y_1(\Delta\omega) * Y_2(\Delta\omega) + Y_1(-\Delta\omega) * Y_2(-\Delta\omega) \\
 &= 2 \operatorname{Re}[Y_1(\Delta\omega) * Y_2(\Delta\omega)] = 0, \\
 & X_1(\Delta\omega) * Y_2(\Delta\omega) - X_1^*(\Delta\omega) * Y_2^*(\Delta\omega) \\
 &= 2 \operatorname{Im}[X_1(\Delta\omega) * Y_2(\Delta\omega)] = 0, \\
 & Y_1(\Delta\omega) * X_2(\Delta\omega) - Y_1^*(\Delta\omega) * X_2^*(\Delta\omega) \\
 &= 2 \operatorname{Im}[Y_1(\Delta\omega) * X_2(\Delta\omega)] = 0.
 \end{aligned} \tag{6.A.3}$$

We analyze the orthogonal conditions for different cases.

### 6.A.1 Band-Limited System

*Without Spectrum Overlap.* Assume that signals are band limited as

$$X_i(\omega) = 0 \quad \text{and} \quad Y_i(\omega) = 0 \quad \text{for} \quad |\omega| \geq \omega_{im} \quad (i = 1, 2).$$

Then Equations 6.A.2 and 6.A.3 hold for  $\Delta\omega \geq \omega_{1m} + \omega_{2m}$ . This is a case where the spectra do not overlap between different carrier signals.

*With Spectrum Overlap.* Consider digital signals with symbol duration of  $T$  and band limitation such that

$$X_i(\omega) = 0, Y_i(\omega) = 0 \quad \text{for} \quad |\omega| \geq \frac{2\pi}{T}, \quad (i = 1, 2)$$

and square root of the Nyquist I roll-off characteristics. This is the optimum transmission system under the white Gaussian noise (Section 3.3.6). If the roll-off factor is zero (rectangular transfer function), the spectra never overlap with each other for  $\Delta\omega = 2\pi/T$ , and OFDM is possible for any QAM system. We will show that for the same channel spacing,  $\Delta\omega = 2\pi/T$ , introduction of the nonzero roll-off factor allows spectrum overlapping.

If we let

$$\begin{aligned}
 x_2(t) &= x_1(t + T/2) \\
 y_2(t) &= y_1(t - T/2)
 \end{aligned} \tag{6.A.4}$$

and assume that  $x_i(t)$  and  $y_i(t)$  are even function of  $t$ , then using notation  $\Delta\omega = 2\pi/T$ , we can see that the first and second equations of Equation 6.A.1 are valid as follows.

Denoting  $t' = t + T/4$ ,

$$\begin{aligned}
 & \int_{-\infty}^{\infty} x_1(t) x_2(t) \cos \Delta\omega t dt \\
 &= - \int_{-\infty}^{\infty} x_1(t' + T/4) x_2(t' - T/4) \sin \Delta\omega t' dt' = 0
 \end{aligned}$$

and denoting  $t' = t - T/4$ ,

$$\begin{aligned} & \int_{-\infty}^{\infty} y_1(t)y_2(t)\cos\Delta\omega tdt \\ &= -\int_{-\infty}^{\infty} y_1(t'+T/4)y_2(t'-T/4)\sin\Delta\omega t'dt'=0 \end{aligned}$$

If we let

$$\begin{aligned} x_1(t) &= y_2(t), \\ y_1(t) &= x_2(t), \end{aligned} \tag{6.A.5}$$

we can similarly confirm that the third and fourth equations of Equation 6.A.1 hold.

From Equations 6.A.4 and 6.A.5, we have

$$\begin{aligned} x_2(t) &= y_1(t) = x_1(t+T/2), \\ y_2(t) &= x_1(t) = x_2(t-T/2) \end{aligned}$$

Similarly, if we let

$$\begin{aligned} x_3(t) &= x_2(t-T/2), \\ y_3(t) &= x_3(t+T/2) \end{aligned}$$

interference with the third subcarrier signals is avoided. Thus, this system becomes an offset QAM (Section 5.3.1), where the time offsets between the inphase and quadrature-phase components are changed in every neighboring subcarriers.

Since  $x_i(t)$  and  $y_i(t)$  are band limited within  $|\omega| < 2\pi/T$ , interchannel interference is absent when channel separation  $\Delta\omega$  becomes higher than  $2\pi/T$  to give no spectrum overlapping.  $N$  subcarriers can be orthogonally multiplexed as shown in Figure 6.65. The transmitter and receiver structure is given with Figure 6.66. This system gives no intersymbol interference since pulse waveforms  $x_i(t)$  and  $y_i(t)$  or  $X_i(\omega)$  and  $Y_i(\omega)$  are given to meet the square root of the Nyquist I roll-off characteristic.

## 6.A.2 Nonband-Limited System

Assume that all  $x_i(t)$  and  $y_i(t)$  ( $i=1, 2$ ) are the same, then if  $X_i(\omega)*Y_i(\omega)$  is zero for  $\omega=n\omega_0$  ( $n=\pm 1, \pm 2, \dots$ ),  $N$  channels can be orthogonal frequency-division multiplexed with spacing of  $\omega_0$ , since Equation 6.A.2 holds.

Let

$$G(\omega) = X_i(\omega) * X_i(\omega),$$

and we require that

$$G(n\omega_0) = \begin{cases} C & (n=0) \\ 0 & (n \neq 0) \end{cases}.$$

where  $C$  is a constant.

The above equation can be rewritten as

$$G(\omega) \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0) = C\delta(\omega)$$

Taking the inverse Fourier transform of the above equation, we get

$$g(t) * \frac{1}{\omega_0} \sum_{n=-\infty}^{\infty} \delta(t - nT) = \frac{C}{2\pi} \quad (6.A.6)$$

where

$$g(t) = x_i^2(t) \leftrightarrow G(\omega).$$

Equation 6.A.6 becomes

$$\sum_{n=-\infty}^{\infty} x_i^2(t - nT) = \frac{C}{T} \quad (6.A.7)$$

The waveforms  $x_i(t)$  do not need to be band-limited. The waveforms that satisfy Equation 6.A.7 are given, for example, by

$$x_i(t) = \begin{cases} 1 & (0 \leq t \leq T) \\ 0 & (\text{otherwise}) \end{cases}$$

or

$$x_i(t) = \begin{cases} \cos\left(\frac{\pi}{2T}t\right) & (-T \leq t \leq T) \\ 0 & (\text{otherwise}) \end{cases}$$

The former is a rectangular waveform and the system becomes the OFDM in the narrow sense. The latter one never satisfies the Nyquist I criterion.

For further description of orthogonal frequency-division multiplexed systems, refer to [81].

## REFERENCES

- [1] Comité Consultatif International Télégraphique et Téléphonique (CCITT) V-series Recommendation.
- [2] Debuda R. Coherent demodulation of frequency shift keying with low deviation ratio. IEEE Trans Commun 1972;COM-20:429–435.
- [3] Pasupathy S. Minimum shift keying: a spectrally efficient modulation. IEEE Commun Mag 1979;17:14–22.
- [4] Murota K, Hirade K. Transmission performance of GMSK modulation. Trans IECE Japan 1981;J64-B:1123–1130.

- [5] Ishizuka M, Hirade K. Optimum Gaussian filter and deviated-frequency-locking scheme for coherent detection of MSK. *IEEE Trans Commun* 1980;COM-28:850–857.
- [6] Kawas Kaleh G. A differentially coherent receiver for minimum shift keying. *IEEE J Select Areas Commun* 1989;7:99–106.
- [7] Masamura T, Samejima S, Morihiro Y, Fuketa F. Differential detection of MSK with nonredundant error correction. *IEEE Trans Commun* 1979;COM-27:912–918.
- [8] Simon MK, Wang CC. Differential versus limiter-discriminator detection of narrow-band FM. *IEEE Trans Commun* 1983;COM-31:1227–1234.
- [9] Amoroso F. Pulse and spectrum manipulation in the minimum (frequency) shift keying (MSK) format. *IEEE Trans Commun* 1976;COM-24:381–384.
- [10] Simon MK. A generalization of minimum-shift-keying (MSK)-type signaling based upon input data symbol pulse shaping. *IEEE Trans Commun* 1976;COM-24:845–856.
- [11] Rabzel N, Pasupathy S. Spectral shaping in minimum shift keying (MSK)-type signals. *IEEE Trans Commun* 1978;COM-26:189–195.
- [12] Lender A. The duobinary techniques for high-speed data transmission. *IEEE Trans Commun Electron* 1963;82:214–218.
- [13] de Jager F, Dekker CB. Tamed frequency modulation, a novel method to achieve spectrum economy in digital transmission. *IEEE Trans Commun* 1978;COM-26:534–542.
- [14] Elnoubi S, Gupta SC. Error rate performance of noncoherent detection of duobinary coded MSK and TFM in mobile radio communication systems. *IEEE Trans Veh Technol* 1981;VT-30:62–72.
- [15] El-Tanany MS, Mahmoud SA. Mean-square error optimization of quadrature receivers for CPM with modulation index 1/2. *IEEE J Select Areas Commun* 1987;SAC-5:896–905.
- [16] Dekker CB. The application of tamed frequency modulation to digital transmission via radio. In: *Proceedings of IEEE National Telecommunication Conference*; 1979. p 55.3.1–55.3.7.
- [17] Muilwijk D. Tamed frequency modulation—a bandwidth—saving digital modulation method, suited for mobile radio. *Philips Telecomm Rev* 1979;37:35–49.
- [18] Muilwijk D. Correlative phase shift keying—a class of constant envelope modulation Techniques. *IEEE Trans Commun* 1981;COM-29:226–236.
- [19] Chung K. Generalized tamed frequency modulation and its application for mobile radio communications. *IEEE Trans Veh Technol* 1984;VT-33:103–113.
- [20] Chung K. Discriminator—MLSE detection of a GTFM signal in the presence of fast Rayleigh fading. *IEEE Trans Commun* 1987;COM-35:1374–1376.
- [21] Hirade K, Murota K. A study of modulation for digital mobile telephony. In: *Proceedings of the 29th IEEE Vehicular Technology Conference Record*, Arlington Heights (IL); 1979. p 13–19.
- [22] Murota K, Hirade K. GMSK modulation for digital mobile radio telephony. *IEEE Trans Commun* 1981;COM-29:1044–1050.
- [23] Suzuki H. Error-rate performance of GMSK modulation with differential detection. *IECE Tech Rep* 1979;CS 79–129:23–30.
- [24] Ogose S, Murota K. Experimental studies on differentially-encoded GMSK with differential-detection. *IECE Tech Rep* 1979;CS 79–130:31–38.
- [25] Ogose S, Murota K. Differentially encoded GMSK with 2-bit differential detection. *Trans IECE* 1981;J64-B:248–254.

- [26] Ogose S. Error-rate performance of differentially encoded GMSK with 2-bit differential detection. *Trans IECE* 1982;J65-B:1253–1259.
- [27] Suzuki H. Optimum Gaussian filter differential detection of MSK. *IEEE Trans Commun* 1981;COM-29:916–918.
- [28] Ogose S. Optimum Gaussian filter for MSK with 2-bit differential detection. *Trans IECE Japan* 1983;E-66:459–460.
- [29] Simon MK, Wang CC. Differential detection of Gaussian MSK in a mobile radio environment. *IEEE Trans Veh Technol* 1984;VT-33:302–320.
- [30] Yongacoglu A, Makrakis D, Feher K. Differential detection of FMSK using decision feedback. *IEEE Trans Commun* 1988;COM-36:641–649.
- [31] Hirono M, Miki T, Murota K. Multilevel decision method for band-limited digital FM with limiter-discriminator detection. *IEEE Trans Veh Technol* 1984;VT-33:114–122.
- [32] Ohono K, Adachi F. Half-bit offset decision frequency detection of differentially encoded GMSK signals. *Electron Lett* 1987;23:1311–1312.
- [33] Elnoubi SM. Analysis of GMSK with discriminator detection in mobile radio channels. *IEEE Trans Veh Technol* 1986;VT-35:71–76.
- [34] Elnoubi SM. Predetection filtering on the probability of error of GMSK with discriminator detection in mobile radio channels. *IEEE Trans Veh Technol* 1988;VT-37:104–107.
- [35] Svenson NAB, Sundberg CEW. Performance evaluation of differential and discriminator detection of continuous phase modulation. *IEEE Trans Veh Technol* 1986;VT-35:106–117.
- [36] Murota K. Spectrum efficiency of GMSK land mobile radio. *IEEE Trans Veh Technol* 1985;VT-34:69–75.
- [37] Okai T, Sugiyama F, Asakawa S, Okamura Y. Narrow band constant envelope digital phase modulation system. *IECE Tech Rep* 1979;CS 79–133:55–62.
- [38] Asakawa S, Sugiyama F. A compact spectrum constant envelope digital phase modulation. *IEEE Trans Veh Technol* 1981;VT-3:102–111.
- [39] Muratore F, Palestini V. Features and performance of 12PM3 modulation methods for digital land mobile radio. *IEEE J Select Areas Commun* 1987;SAC-5:906–914.
- [40] Maseng T. Digitally phase modulated (DPM) signals. *IEEE Trans Commun* 1985;COM-33:911–918.
- [41] Aulin T, Sundberg C-EW. Continuous phase modulation—part I: full response signaling. *IEEE Trans Commun* 1981;COM-29:196–209.
- [42] Aulin T, Rydbeck N, Sundberg C-EW. Continuous phase modulation—part II: partial response signaling. *IEEE Trans Commun* 1981;COM-29:210–225.
- [43] Anderson JB, Aulin T, Sundberg CE. *Digital Phase Modulation*. New York: Plenum Press; 1986.
- [44] Miyakawa H, Harashima H, Tanaka Y. A new digital modulation scheme, multimode binary CPFSK. In: *Proceedings of Third International Conference on Digital Satellite Communication*, Kyoto, 1975. p 105–112.
- [45] Takagi K, Yamamoto B. Performance of narrow band digital transmission system with a duobinary filter. *Proc IECE Natl Convent* 1981;2169.
- [46] Takagi K, Yamamoto B. Narrow band digital FM scheme using duobinary filter. In: *Proceedings of the IEEE National Telecommunication Conference*; 1981. p B.8.5.1–B8.5.5.

- [47] Akaiwa Y, Takase I, Ikoma M, Saegusa N. An FM modulation-demodulation scheme for narrow-band digital communication. IECE Tech Rep 1979;CS 79-132:47-54.
- [48] Akaiwa Y, Takase I, Kojima S, Ikoma M, Saegusa N. Performance of baseband-bandlimited multilevel FM with discriminator detection for digital mobile telephony. Trans IECE 1981;E64:463-469.
- [49] Akaiwa Y, Kojima S. Performance of baseband bandlimited 4-level FM with coherent and differential detection. Proc IECE Natl Convent Commun 1980;475.
- [50] Honma K, Murata E, Tatsuzawa Y. A study of digital mobile radio communication method using PLL. IECE Tech Rep 1979;CS 79-134.
- [51] Honma K, Murata E, Rikou Y. On a method of constant envelope modulation for digital mobile radio communication. In: Proceedings of the IEEE International Conference on Communication, Seattle; 1980. p 24.1.1-24.1.5.
- [52] Akaiwa Y, Okamoto E. An analysis of error rates for Nyquist- and partial response-baseband-filtered digital FM with discriminator detection. Trans IECE 1983;J66-B: 534-541.
- [53] Jenks FG, Morgan PD, Warren CS. Use of four level phase modulation for digital mobile radio. IEEE Trans Electromag Compat 1972;EMC-14:113-128.
- [54] Y. Akaiwa and Y. Nagata, A linear modulation method for digital mobile radio communication. In: Proceedings of 1985 National Convention, IECE, No. 2384. Nagata Y, Akaiwa Y. Characteristics of a linear modulation method for digital mobile radio communications. *ibid.*, No. 2385.
- [55] Akaiwa Y, Nagata Y. Highly efficient digital mobile radio communications with a linear modulation method. IEEE J Select Area Commun 1987;SAC-5:890-895.
- [56] Tarralo JA, Zysman GI. Modulation techniques for digital cellular systems. In: Proceedings of the IEEE Vehicular, Technology Conference, 1988. p 245-248.
- [57] Liu C, Feher K. Noncoherent detection of  $\pi/4$ -QPSK systems in a CCI-AWGN combined interference environment. In: Proceedings of IEEE Vehicular, Technology Conference, 1989. p 83-94.
- [58] Adachi F, Mori M, Ooi T. Radio channel structure for QPSK digital mobile radio. In: Proceedings of IEEE Vehicular Technology Conference, 1989. p 220-223.
- [59] Raith K, Hedberg B, Larsson G, Kahre R. Performance of a digital cellular experiment test bed. In: Proceedings of the IEEE Vehicular Technology Conference, 1989. p 175-177.
- [60] Tomisato S, Suzuki H. Envelope controlled digital modulation improving power efficiency of transmitter amplification—an application to trellis-coded 8PSK for mobile radio. Trans IEICE 1992;J75-B-II:918-928.
- [61] Martin PM, Bateman A, McGeehan JP, Marvil JD. The implementation of a 16-QAM mobile data system using TTIB-based fading correction techniques. In: Proceedings of the IEEE Vehicular, Technology Conference, Philadelphia: Institute of Electrical and Electronics Engineers (IEEE); 1988. p 71-76.
- [62] Sampei S, Sunaga T. Rayleigh fading compensation method for 16-QAM in digital land mobile radio channels. In: Proceedings of the IEEE Vehicular, Technology Conference; San Francisco (CA); 1989. p 640-646.
- [63] Yokoyama M. BPSK system with sounder to combat Rayleigh fading in mobile radio communications. IEEE Trans Veh, Technol 1985;VT-34:35-40.
- [64] Simon MK. Dual-pilot tone calibration technique. IEEE Trans Veh Technol 1986;VT-35: 63-70.

- [65] N. Kinoshita S. Sampei, E Moriyama, H Sasaoka, Y. Kamio, K. Miya, K. Hiramatsu, K. Inogai, K. Homma, Field experiments on 16QAM/TDMA and trellis coded 16QAM/TDMA systems for digital land mobile radio communications, IEICE Trans Commun, E77-B, 911–920 (1994).
- [66] Dixon RC. *Speed Spectrum System*. 2nd ed. New York: Wiley-Interscience; 1984.
- [67] Simon MK, Omura JK, Scholtz RA, Levitt BK. *Spread Spectrum Communications*. Volume I, II, III, Rockville, MD: Computer Science Press; 1985.
- [68] Chang RW. Synthesis of band-limited orthogonal signals for multichannel data transmission. Bell Syst Tech J 1966;45:1775–1796.
- [69] Shmidman DA. A generalized Nyquist criterion and an optimum linear receiver for a pulse modulation system. Bell Syst Tech J 1967;46:2163–2177.
- [70] Saltberg BR. Performance of an efficient parallel data transmission system. IEEE Trans Commun Technol 1967;COM-15:805–811.
- [71] Hirotsaki B. An orthogonally multiplexed QAM system using the discrete Fourier transform. IEEE Trans Commun 1981;COM-29:982–989.
- [72] Horikoshi J. Error performance considerations of  $\pi/2$ -TFSK under the multipath interfering environment. Trans IECE 1984;E67:40–46.
- [73] Yoshida S, Ariyavisitakul S, Ikegami F, Takeuchi T. A novel anti-multipath modulation technique DSK. In: Proceedings of the IEEE Globecom, 1985. p 36.4.1–36.4.5.
- [74] Yoshida S, Ariyavisitakul S, Ikegami F, Takeuchi T. A power-efficient linear digital modulation and its application to anti-multipath modulation PSK-RZ scheme. In: Proceedings of the IEEE Vehicular Technology Conference, Tampa, June 1987. p 66–71.
- [75] Yoshida S, Ikegami F. Anti-multipath modulation technique-Manchester-coded DPSK and its generalization. IEICE Tech Rep 1986;CS 86–47.
- [76] Takai H. BER performance of anti-multipath modulation PSK-VP and its optimum phase-waveform. In: Proceedings of the IEEE 40th Vehicular Technology Conference, 1990. p 412–419.
- [77] Alard M, Lassalle R. Principles of modulation and channel coding for digital broadcasting for mobile receivers. EBU Rev Tech 1987;224:168–190.
- [78] Hirotsaki B, Aoyagi H. A highly efficient HF modem with adaptive fading control algorithm. In: Proceedings of the IEEE Globecom, 1984. p 48.3.1–48.3.5.
- [79] Cimini LJ. Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing. IEEE Trans Commun 1985;COM-33 (7):665–675.
- [80] Weinstein SB. The history of orthogonal frequency-division multiplexing. IEEE Commun Mag 2009;26–35.
- [81] Hanzo L, Munster M, Choi BJ, Keller T. *OFDM and MC-CDMA*. New York: IEEE Press; 2003.
- [82] Maruta R, Tomozawa A. An improved method for digital SSB-FDM modulation and demodulation. IEEE Trans Commun 1978;COM-26:720–725.
- [83] Myung H, Lim J, Goodman D. single carrier FDMA for uplink wireless transmission. IEEE Veh Mag 2006;30–38.

This page intentionally left blank



## OTHER TOPICS IN DIGITAL MOBILE RADIO TRANSMISSION

This chapter introduces other important technologies. Diversity transmission system, multi-input multi-output (MIMO) systems, adaptive automatic equalizers, error control techniques, trellis modulation, adaptive interference cancellation, and voice coding are described.

### 7.1 DIVERSITY TRANSMISSION SYSTEM

Diversity techniques mitigate fading by using multiple received signals that experience different fading conditions. This system is based on the fact that it is unlikely that all signals experience simultaneously bad fading conditions. A fading channel causes three kinds of transmission errors: (1) thermal noise error due to level fading, (2) irreducible error due to random FM effects, and (3) errors due to waveform distortion caused by frequency-selective fading. The term irreducible errors means that the error rates cannot be reduced even if average signal to noise power ratio (SNR) is arbitrarily increased. A diversity system is effective for reducing the three kinds of errors. The effect of a diversity system on the thermal noise error is easily understood since the signal level dropping probability is diminished with a diversity system. The random FM effect and the waveform distortion become significant when the signal level drops. Thus, we can understand that the random FM effect and the waveform distortion caused by the fading channel are mitigated in the diversity system by preventing the signal level from dropping to a low value.

A detailed description of diversity systems is beyond the scope of this book. Instead, a brief introductory discussion is given followed by a description of the error rate analysis and some diversity systems used for the digital mobile radio communications. For comprehensive treatments, refer to [1, 2].

Many different diversity systems are known: space diversity with multiple antennas, polarization diversity using differently polarized waves, frequency diversity with multiple frequencies, time diversity by transmission of the same signal in different times, and angle diversity using directive antennas aimed in different directions.

With respect to the signal combining methods, we have maximal ratio combining, equal gain combining, and selection (switching) combining. These are further classified into RF, IF, and postdemodulation combining. RF combining and IF (or pre-demodulation) combining require synchronization so that the modulated signals are in phase with each other. Without synchronization, the diversity effect is lost for maximum ratio or equal gain combining and switching noise appears for the selection diversity. This problem is absent with postdemodulation combining; however, the technique requires two or more receivers.

Another category of classification of diversity communications includes transmitter diversity and receiver diversity, using multiple transmitters and receivers, respectively.

Space diversity is classified into micro-diversity and macro- (or site) diversity, according to whether antennas are closely spaced of the order of a wavelength or separated wide enough to cope with the topographical conditions (e.g., buildings, road and terrain). The former is effective for fast fading where signal fades in a distance of the order of a half-wavelength. The latter is effective for shadowing, where the signal fades due to the topographical obstructions (Section 4.2). The micro-diversity can be applied to a terminal where antenna elements are placed closely in a small area of the terminal. If the same signal is transmitted from the antenna elements at the same time without controlling the antenna weights, a directional pattern is generated which never gives the diversity effect.

To get a desired directional pattern which gives the diversity effect, we should control the antenna weights so as to make the received signals transmitted from the antenna elements are combined in the same phase depending on the propagation factors of each antenna paths to a receiver antenna (Section 7.2.1). Some approaches are known, which never requires such a treatment at the transmitter. One is the cyclic delay diversity, where a transmit signal is divided into blocks, and  $N$  signal blocks are made by cyclic shifts within a signal block with different shifts corresponding to each antenna element of  $N$ . A cyclic shifted signal block is sent from the antenna element at the same time. A directional antenna pattern is never generated since the antenna elements transmit different signal at the same time. This is a kind of the time diversity. Received signal is equalized with a decision feedback equalizer (Section 7.3.3) or a frequency-domain equalizer (Section 7.3.7) similar to a time diversity system. By using the minimum mean square criteria for the equalizers, we get the space- and time-diversity effect: even when the channels show a high correlation during a time period where the signal is sent, we have the space diversity effect if the correlation is between channels from different transmit antenna to the receiver antenna.

The other method is known as the time-space code, where signal is transmitted in time shift of a symbol period. This system is discussed in Section 7.2.2.

All of the above-mentioned systems are introduced intentionally to mitigate fading; they can be called explicit diversity. The “implicit diversity” effect is obtained unintentionally when an automatic equalizer is used in frequency-selective fading channels. In a frequency-selective fading channel, signals are received subject to different fading conditions and different propagation delays depending on the details of multipath propagation. An automatic equalizer (Section 7.3), introduced to equalize the channel distortion due to the multipath fading, uses time diversity in the channel equalizing process by combining signals from different paths. The same effect is obtained by a spread-spectrum RAKE receiver (Section 6.4).

### 7.1.1 Probability Density Function of SNR for Diversity System

First, we discuss error rates due to the thermal noise in fading channels with diversity systems. The average error rate due to the thermal noise is given as

$$\langle P_e \rangle = \int_0^{\infty} p_e(\gamma) p(\gamma) d\gamma \quad (7.1)$$

Where  $p_e(\gamma)$  is the error rate for a given SNR (signal to noise power ratio), denoted by  $\gamma$ , and  $p(\gamma)$  is the probability density function of  $\gamma$ . Without a diversity system,  $p(\gamma)$  is assumed here to take an exponential distribution (Eq. 4.11a) that is given by the Rayleigh distribution for signal envelope.

Assume the same mean signal power and mean noise power for all branches. Then mean SNR is the same for all branches. The local mean SNR for the  $i$ th branch,  $\gamma_i$ , is given as

$$p(\gamma_i) = \frac{1}{\gamma_0} e^{-\frac{\gamma_i}{\gamma_0}} \quad (0 \leq \gamma_i \leq \infty)$$

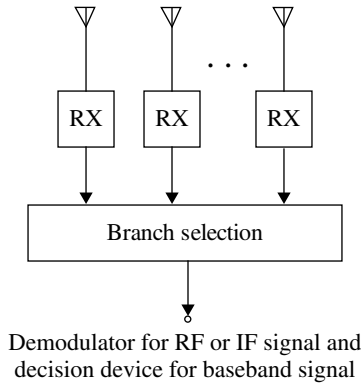
where  $\gamma_0$  is the average SNR.

The probability that  $\gamma_i$  takes values less than  $\gamma$  is

$$P(\gamma_i \leq \gamma) = \int_0^{\gamma} p(\gamma_i) d\gamma_i = 1 - e^{-\gamma/\gamma_0}$$

**Selection Diversity.** Let us consider an  $M$  branch space diversity system with the selection combining method shown in Figure 7.1. Assume that the received signal levels are independent of each other. The branch with the highest  $\gamma$  is selected at each instant. Denote the highest  $\gamma_i$  in the selected branch as  $\gamma$  at an instant. The other branches have  $\gamma_i$  less than  $\gamma$ . Thus, the probability that this situation occurs, that is, the selection diversity system output SNR yielding  $\gamma$ , becomes

$$P_M(\gamma_1, \gamma_2, \dots, \gamma_M \leq \gamma) = [1 - e^{-\gamma/\gamma_0}]^M$$



**FIGURE 7.1** Space diversity selection combining system.

Differentiating the above equation with respect to  $\gamma$ , we have the probability density function as

$$p(\gamma) = \frac{d}{d\gamma} \left[ 1 - e^{-\gamma/\gamma_0} \right]^M \quad (7.2a)$$

$$= \frac{M}{\gamma_0} e^{-\gamma/\gamma_0} \left[ 1 - e^{-\gamma/\gamma_0} \right]^{M-1} \quad (7.2b)$$

$$\approx M \frac{\gamma^{M-1}}{\gamma_0^M} (\gamma \ll \gamma_0) \quad (7.2c)$$

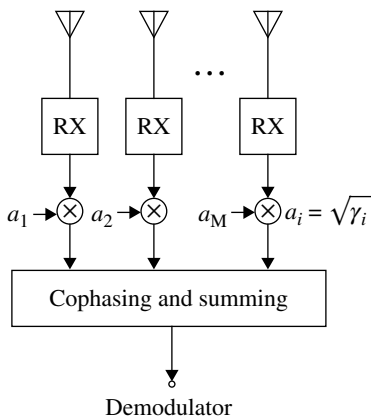
**Maximal Ratio Combining.** The maximal ratio combining (MRC) system is shown in Figure 7.2. Received signals at each branch are weighted by a factor of each branch SNR and coherently summed. This system gives the best error rate performance. From the results in Section 7.2.1, the probability density function of  $\gamma$  at the output of the combiner is given by  $M$ -times convolution (Section 2.2.8) as [1]

$$p(\gamma) = \frac{1}{(M-1)!} \frac{\gamma^{M-1}}{\gamma_0^M} e^{-\gamma/\gamma_0} \quad (7.3a)$$

$$\approx \frac{1}{(M-1)!} \frac{\gamma^{M-1}}{\gamma_0^M} (\gamma \ll \gamma_0) \quad (7.3b)$$

The ratio between Equations 7.3b and 7.2c becomes  $1/M!$ . This means that the average error rate with the maximal ratio combiner is  $1/M!$  times smaller than that for selection combining.

**Equal Gain Combining.** The equal gain combining system is given by taking the same absolute value for weighting factors ( $|a_i| = a_0$  for all  $i$ ) in the MRC system.



**FIGURE 7.2** Maximal ratio combining system.

Unfortunately, the probability density function for the equal gain combining is not obtained in closed form. Approximate expression for  $\gamma \ll \gamma_0$  is given [1] as

$$p(\gamma) \approx \frac{2^{M-1} M^M}{(2M-1)!} \frac{\gamma^{M-1}}{\gamma_0^M} \quad (7.4)$$

For  $M \gg 1$ , MRC has advantage by 1.3 dB over equal gain combining. We can guess intuitively that the degree of the effect of equal gain combining on SNR distribution improvement is between those for maximal ratio combining and selection combining.

**Effects of Branch Correlation** [1, 2]. If the branch SNRs are correlated, the diversity effect is diminished. For an extreme case of complete correlation, we have no diversity effect. Here, the results of analyses for a two-branch system ( $M=2$ ) are cited as follows.

*Selection Diversity.* The distribution function is given as

$$P_2(\gamma) = 1 - e^{-\gamma/\gamma_0} [1 - Q(a, b) + Q(b, a)] \quad (7.5)$$

where

$$Q(a, b) = \int_b^\infty \frac{1}{2} e^{-\frac{1}{2}(a^2 + x^2)} I_0(ax) x dx \quad \left( b = \sqrt{2\gamma/\gamma_0 (1 - \rho^2)}, a = b\rho \right)$$

where  $I_0(x)$  is the zeroth-order modified Bessel function of the first kind and  $\rho$  denotes the magnitude of the complex cross covariance of fading Gaussian processes. It is shown that the normalized envelope covariance of the two-branch signals is very near  $\rho^2$ . The average error rate is approximately determined by probabilities for  $\gamma \ll \gamma_0$ . For  $\gamma \ll \gamma_0$ , Equation 7.5 is approximated as

$$P_2(\gamma) \approx \frac{\gamma^2}{\gamma_0^2 (1 - \rho^2)}$$

The probability density function is given as

$$p_2(\gamma) = \frac{d}{d\gamma} P_2(\gamma) \approx \frac{2\gamma}{\gamma_0^2(1-\rho^2)} \quad (7.6)$$

**MRC.** The distribution function is given as

$$P_2(\gamma) = 1 - \frac{1}{2\rho} \left[ (1+\rho)e^{-\gamma/\gamma_0(1+\rho)} - (1-\rho)e^{-\gamma/\gamma_0(1-\rho)} \right]$$

For  $\gamma \ll \gamma_0$ ,

$$P_2(\gamma) \approx \frac{\gamma^2}{2\gamma_0^2(1-\rho^2)}$$

The probability density function becomes

$$p_2(\gamma) = \frac{1}{2\rho\gamma_0} \left[ e^{-\gamma/\gamma_0(1+\rho)} - e^{-\gamma/\gamma_0(1-\rho)} \right] \quad (7.7a)$$

For  $\gamma \ll \gamma_0$ ,

$$p_2(\gamma) \approx \frac{\gamma}{\gamma_0^2(1-\rho^2)} \quad (7.7b)$$

Comparing Equations 7.6 and 7.7b and considering Equation 7.1, we can see that selection diversity has an average (approximate) error rate twice that of the MRC system. The effect of the correlation between the two-branch signals is reflected in the term  $\gamma_0^2(1-\rho^2)$  in Equations 7.6 and 7.7b. Hence, the effect of the correlation is equivalent to reducing the average SNR by  $\sqrt{1-\rho^2}$ . For example, SNR is reduced by 0.02, 0.14, and 0.62 dB for  $\rho=0.1$ , 0.25, and 0.5, respectively.

### 7.1.2 Average Error Rate for Diversity Systems

**7.1.2.1 Error Rates due to Additive Gaussian Noise** Error rates for a given bit energy to noise power density ratio  $E_b/N_0 (\equiv \gamma)$  are given in the following. For BPSK and QPSK with coherent detection (Eqs. 5.18a and 5.21), we have

$$P_e(\gamma) = \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) \quad (7.8)$$

where the parameter  $\alpha$  is introduced for convenience as a degradation factor. For ASK and FSK with noncoherent detection and BPSK with differential detection (Eqs. 5.30 and 5.31), we have

$$P_e(\gamma) = \frac{1}{2} e^{-\alpha\gamma} \quad (7.9)$$

where  $\alpha=1/2$  for ASK and FSK and  $\alpha=1$  for BPSK with ideal performance.

The average error rate is calculated as

$$\langle P_e \rangle = \int_0^\infty \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) p(\gamma) d\gamma \quad (\text{coherent detection})$$

and

$$\langle P_e \rangle = \int_0^\infty \frac{1}{2} e^{-\alpha\gamma} p(\gamma) d\gamma \quad (\text{noncoherent detection})$$

**MRC.** From Equations 7.1 and 7.3a, the average error rate is given by

$$\langle P_e \rangle = \int_0^\infty P_e(\gamma) \frac{1}{(M-1)!} \frac{\gamma^{M-1}}{\gamma_0^M} e^{-\gamma/\gamma_0} d\gamma$$

For noncoherent detection, using Equation 7.9 and integrating the above equation by part for  $M=1, 2, \dots$ , we have

$$\langle P_e \rangle = \frac{1}{2} \frac{1}{(1 + \alpha\gamma_0)^M} \quad (\text{noncoherent detection}) \quad (7.10)$$

For low error rates,

$$\langle P_e \rangle = \frac{1}{2} \frac{1}{(\alpha\gamma_0)^M} \quad (\alpha\gamma_0 \gg 1) \quad (7.11)$$

The same result as this approximate formula can be obtained using the approximate probability density function for low  $\gamma$  as

$$\langle P_e \rangle \approx \int_0^\infty \frac{1}{2} e^{-\alpha\gamma} \frac{1}{(M-1)!} \frac{\gamma^{M-1}}{\gamma_0^M} d\gamma \quad (\gamma \ll \gamma_0) \quad (7.12a)$$

$$= \frac{1}{2} \frac{1}{(\alpha\gamma_0)^M} \quad (7.12b)$$

For coherent detection,

$$\langle P_e \rangle = \int_0^\infty \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) \frac{1}{(M-1)!} \frac{\gamma^{M-1}}{\gamma_0^M} e^{-\gamma/\gamma_0} d\gamma \quad (7.13a)$$

For  $M \geq 2$ , we have (Appendix 7.A)

$$\langle P_e \rangle = \frac{1}{2} - \frac{1}{2} \frac{1}{\sqrt{1 + 1/(\alpha\gamma_0)}} \left[ 1 + \sum_{m=1}^{M-1} \frac{(2m-1)!! / (2m)!!}{(1 + \alpha\gamma_0)^m} \right] \quad (7.13b)$$

(coherent detection)

For  $M=2$ ,

$$\langle P_e \rangle = \frac{1}{2} - \frac{1}{2} \frac{1}{\sqrt{1 + \frac{1}{\alpha\gamma_0}}} \left( 1 + \frac{1}{2} \frac{1}{1 + \alpha\gamma_0} \right) \quad (7.14a)$$

$$\approx \frac{3}{16} \frac{1}{(\alpha\gamma_0)^2} \quad (\alpha\gamma_0 \gg 1) \quad (7.14b)$$

If we use the approximate probability density function, we have (Appendix 7.B)

$$\langle P_e \rangle \approx \int_0^1 \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) \frac{1}{(M-1)!} \frac{\gamma^{M-1}}{\gamma_0^M} d\gamma \quad (7.15a)$$

$$= \frac{1}{2} \frac{(2M-1)!!}{(2M)!!} \frac{1}{(\alpha\gamma_0)^M} \quad (7.15b)$$

For  $M=2$ , Equation 7.15b yields the same result as Equation 7.14b.

**Selection Diversity.** Using Equations 7.1 and 7.2a, we have

$$\langle P_e \rangle = \int_0^\infty P_e(\gamma) \frac{d}{d\gamma} (1 - e^{-\gamma/\gamma_0})^M d\gamma$$

Integrating by part,  $\langle P_e \rangle$  becomes

$$\langle P_e \rangle = \left[ P_e(\gamma) (1 - e^{-\gamma/\gamma_0})^M \right]_0^\infty - \int_0^\infty \left\{ \frac{d}{d\gamma} P_e(\gamma) \right\} (1 - e^{-\gamma/\gamma_0})^M d\gamma \quad (7.16)$$

Considering  $P_e(\infty) = 0$ , and expanding as  $(1 - e^{-\gamma/\gamma_0})^M$

$$(1 - e^{-\gamma/\gamma_0})^M = \sum_{k=0}^M \binom{M}{k} (-1)^k e^{-k\gamma/\gamma_0}$$

we have

$$\langle P_e \rangle = \sum_{k=0}^M \binom{M}{k} (-1)^k \int_0^\infty \left\{ \frac{d}{d\gamma} P_e(\gamma) \right\} e^{-k\gamma/\gamma_0} d\gamma \quad (7.17)$$

where  $\binom{M}{k} = \frac{M!}{(M-k)!k!}$

For noncoherent detection, using Equations 7.9 and 7.17, we have

$$\begin{aligned} \langle P_e \rangle &= \frac{1}{2} \sum_{k=0}^M \binom{M}{k} (-1)^k \int_0^\infty \alpha e^{-(\alpha+k/\gamma_0)\gamma} d\gamma \\ &= \frac{1}{2} \sum_{k=0}^M \binom{M}{k} (-1)^k \frac{1}{1 + k/(\alpha\gamma_0)} \quad (\text{noncoherent detection}) \end{aligned} \quad (7.18)$$



For  $M=2$ ,

$$\langle P_e \rangle = \frac{1}{2} - \frac{1}{1 + 1/(\alpha\gamma_0)} + \frac{1}{2} \frac{1}{1 + 2/(\alpha\gamma_0)} \quad (7.19a)$$

$$\approx \frac{1}{(\alpha\gamma_0)^2} \quad (\alpha\gamma_0 \gg 1) \quad (7.19b)$$

If we use the approximate probability distribution (Eq. 7.2c),

$$\langle P_e \rangle \approx \int_0^\infty \frac{1}{2} e^{-\alpha\gamma} M \frac{\gamma^{M-1}}{\gamma_0^M} d\gamma$$

Comparing this with Equation 7.12a, we have

$$\langle P_e \rangle \approx \frac{1}{2} \frac{M!}{(\alpha\gamma_0)^M}$$

This result for  $M=2$  is the same as Equation 7.19b.

For coherent detection,

$$\frac{d}{d\gamma} P_e(\gamma) = -\frac{1}{2\sqrt{\pi}} \sqrt{\frac{\alpha}{\gamma}} e^{-\alpha\gamma}$$

Inserting this into Equation 7.17 and using the relation

$$\int_0^\infty \frac{1}{\sqrt{\gamma}} e^{-\beta\gamma} d\gamma = \sqrt{\frac{\pi}{\beta}}$$

we have

$$\langle P_e \rangle = \frac{1}{2} \sum_{k=0}^M \binom{M}{k} (-1)^k \frac{1}{\sqrt{1 + k/(\alpha\gamma_0)}} \quad (\text{coherent detection}) \quad (7.20)$$

For  $M=2$ ,

$$\langle P_e \rangle = \frac{1}{2} - \frac{1}{\sqrt{1 + 1/(\alpha\gamma_0)}} + \frac{1}{2} \frac{1}{\sqrt{1 + 2/(\alpha\gamma_0)}} \quad (7.21a)$$

$$\approx \frac{3}{8} \frac{1}{(\alpha\gamma_0)^2} \quad (7.21b)$$

If we use the approximate probability distribution function, then

$$\langle P_e \rangle \approx \int_0^\infty \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) M \frac{\gamma^{M-1}}{\gamma_0^M} d\gamma \quad (\gamma \ll \gamma_0)$$

Comparing this with Equation 7.15a, we get

$$\langle P_e \rangle \approx \frac{M! (2M-1)!!}{2 (2M)!!} \frac{1}{(\alpha\gamma_0)^M} \quad (\text{coherent detection})$$

For  $M=2$ , we have the same result as Equation 7.21b.

**Equal Gain Combining.** Since no exact expression for the probability density function  $\gamma$  is known, we use the approximate probability density function given by Equation 7.4. Comparing Equation 7.4 with Equation 7.3b and using Equations 7.12b and 7.15b, we get

$$\langle P_e \rangle \approx \begin{cases} \frac{1}{2} \frac{(2M)^M M!}{(2M)!!} \frac{1}{(\alpha\gamma_0)^M} & (\text{noncoherent detection}) \\ \frac{1}{2} \frac{(2M)^M M! (2M-1)!!}{(2M)!(2M)!!} \frac{1}{(\alpha\gamma_0)^M} & (\text{coherent detection}) \end{cases} \quad (7.22)$$

For  $M=2$ ,

$$\langle P_e \rangle \approx \begin{cases} \frac{2}{3} \frac{1}{(\alpha\gamma_0)^2} & (\text{noncoherent detection}) \\ \frac{1}{4} \frac{1}{(\alpha\gamma_0)^2} & (\text{coherent detection}) \end{cases}$$

Using Equations 7.10, 7.13a, 7.18, 7.20, and 7.22, bit error rates are calculated as shown in Figure 7.3a and b.

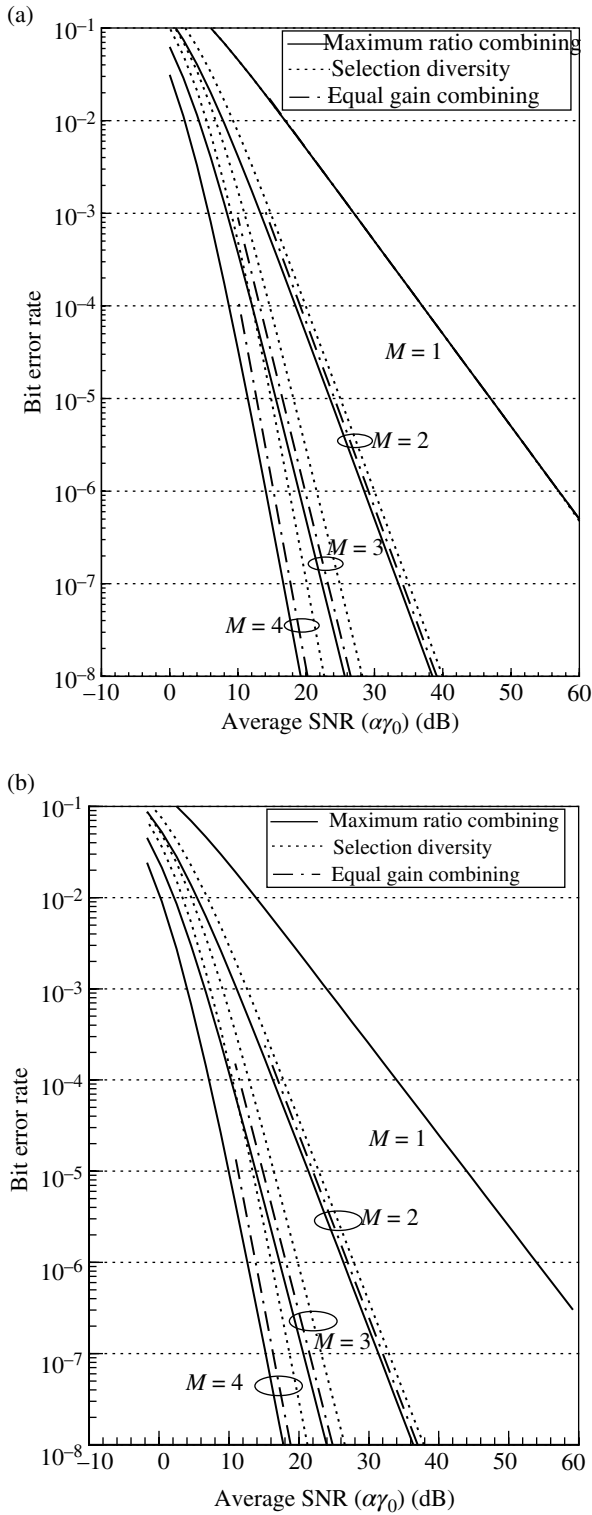
The postdetection combining diversity system for QPSK and  $\pi/4$  shifted QPSK with differential detection is extensively analyzed [3] (Section 5.5.5). When square root raised cosine roll-off Nyquist filters are used for the transmit and receive filters, a simple result is given as [3]

$$\langle P \rangle_e \approx K_M \left( \frac{1}{\gamma_0} \right)^M \quad (7.23)$$

where  $\gamma_0$  is the bit energy to noise power density ratio and  $K_M$  is given as

$$K_M = \begin{cases} \frac{(2M-1)!!}{2} & (\text{selection combining}) \\ \frac{1}{2} \frac{M^M}{M!} & (\text{equal gain combining}) \\ \frac{1}{2} \frac{(2M-1)!!}{M!} & (\text{maximal ratio combining}) \end{cases}$$

**7.1.2.2 Irreducible Error Rates due to Random FM Effect** In contrast to the error rate analysis for the thermal noise effects, it is difficult to evaluate the effects of



**FIGURE 7.3** Error rates for diversity systems. (a) noncoherent detection and (b) coherent detection.

diversity on the random FM effect and waveform distortion resulting from fading channels. This is because the various parameters of systems such as modulation/demodulation method, pulse waveforms, and the diversity system must be considered.

The two-branch postdetection combining system for BPSK with differential detection is analyzed in [2]. The result is cited here as

$$\langle P_e \rangle = \left( \frac{1 + \gamma_0 [1 - J_0(2\pi f_m T)]}{2(1 + \gamma_0)} \right)^2 \left( \frac{2(\gamma_0 + 1) + \gamma_0 J_0(2\pi f_m T)}{1 + \gamma_0} \right)$$

where  $f_m$  is the maximum Doppler frequency and  $T$  is the symbol duration. When we let the average signal to noise power ratio  $\gamma_0$  to be infinite, the irreducible error rate is given as

$$\langle P_e \rangle_{\text{ir}} \approx \frac{1}{4} [1 - J_0(2\pi f_m T)]^2 [2 + J_0(2\pi f_m T)]$$

This is further simplified for  $f_m T \ll 1$ :

$$\langle P_e \rangle_{\text{ir}} \approx \frac{3}{4} (\pi f_m T)^4 \quad (7.24)$$

where  $J_0(x) \approx 1 - (x/2)^2$  ( $x \ll 1$ ) is used.

If this result is compared with that for the single branch system given as

$$\langle P_e \rangle_{\text{ir}} \approx \frac{1}{2} (\pi f_m T)^2 \quad (7.25)$$

we can see that the irreducible error rate, which is the second order of a small value  $f_m T$ , is reduced to fourth order of a small value  $f_m T$  with the two-branch diversity.

If we assume slow fading ( $f_m T \rightarrow 0$ ), then the error rate corresponds to the thermal noise effect alone. The error rates are improved with diversity as  $\langle P_e \rangle_{\text{th}} \propto \gamma_0^{-M}$ , where  $M$  is the number of branches.

Results for a postdetection combining diversity system for QPSK and  $\pi/4$  shifted QPSK with differential detector are given [3] as

$$\langle P_e \rangle_{\text{ir}} \approx K_M (2\pi f_{\text{rms}} T)^{2M}$$

Also for this case, the irreducible error rates are reduced as the  $M$ th order of  $(f_{\text{rms}} T)^2$  similarly to Equations 7.24 and 7.25.

The irreducible error rate of a postdetection diversity system for digital FM with frequency detection is given as [2]

$$\langle P_e \rangle_{\text{ir}} \approx \frac{(2M-1)!}{(M-1)!^2} \left( \frac{f_m}{2\sqrt{2}\Delta f} \right)^{2M}$$

where  $\Delta f$  is the peak frequency deviation. Using the modulation index  $m$ ,  $\Delta f$  is expressed as  $m/2T$ . Thus,  $\langle P_e \rangle_{\text{ir}}$  is proportional to  $[(f_{\text{rms}} T)^2]^M$  for a given modulation index.

**7.1.2.3 Error Rates due to Frequency Selective Fading** For the raised cosine roll-off filtered ( $\pi/4$  shifted) QPSK, error rates under frequency-selective fading are given in Ref. [3].

**7.1.2.4 Error Rates due to Co-Channel Interference** For the raised cosine roll-off filtered ( $\pi/4$  shifted) QPSK, approximate error rates are given in Ref. [3].

### 7.1.3 Multiple Transmitter Diversity System

In this system, a signal is transmitted simultaneously from multiple base stations that are located far away from each other. Thus, it is a macrodiversity system. This technique is used in some commercial paging systems. For paging systems, a crucial issue is the expansion of the service area with low transmit power. For this purpose, a multiple base station transmit system is appropriate. In addition, since it is a macrodiversity system, it can mitigate the effects of shadowing as well as of fast fading in areas where the zones of the base station overlap.

However, attention must be paid to interference between the signals from different base stations. At least bit timings are synchronized with each other. Some digital modulation techniques have been proposed to avoid interference effects and to obtain diversity effects [5–7]. Digital FM is exclusively used for paging systems, since transmit power efficiency is the main concern for the high power transmitter and since the receivers must be simple and small. The multiple base station simultaneous transmission systems proposed so far use different baseband signals between the transmitters. The limiter–discriminator detector is used for the receiver. The integrate-and-dump filter is also used since the transmit baseband signal is the NRZ rectangular pulse (Section 3.3.3).

The diversity effect is analyzed by Adachi [7] as follows. Consider a two-transmitter system. We denote the transmitted signals as

$$\begin{aligned} u_1(t) &= \cos\{\omega_0 t + \varphi_1(t)\} \\ u_2(t) &= \cos\{\omega_0 t + \varphi_2(t)\} \end{aligned}$$

where  $\omega_0$  is the carrier frequency and  $\varphi_i$  ( $i = 1, 2$ ) denotes the baseband phase signal. Assuming the Rayleigh fading, the received signals are expressed as

$$v_i(t) = R_i(t) \cos\{\omega_0 t + \varphi_i(t) + \theta_i(t)\} \quad (i = 1, 2)$$

where  $R_i(t)$  is the envelope, which is Rayleigh distributed, and the phase  $\theta_i(t)$  has a uniform distribution between 0 and  $2\pi$  (Section 4.3).

At a receiver, the limiter–discriminator output signal  $v(t)$  relative to  $\omega_0$  is given, when  $R_1(t) > R_2(t)$ , as

$$v(t) = \dot{\varphi}(t) + \dot{\theta}_1(t) - \frac{d}{dt} \left[ \tan^{-1} \frac{\alpha(t) \sin \psi(t)}{1 + \alpha(t) \cos \psi(t)} \right] \quad (7.26)$$

where the dot denotes the time derivative and

$$\alpha(t) = \frac{R_2(t)}{R_1(t)} \quad (7.27)$$

$$\psi(t) = \varphi_1(t) - \varphi_2(t) + \theta_1(t) - \theta_2(t) \quad (7.28)$$

where  $\psi(t)$  is the beat frequency signal. When  $R_1(t) < R_2(t)$ , we should interchange the subscript in Equations 7.26–7.28. The third term on the right-hand side of Equation 7.26 represents the interference between the two received signals. This term must disappear for no interference between the two signals. Since the fading is relatively slow compared with the bit rate  $1/T$ ,  $R_i(t)$  and  $\theta_i(t)$  remain constant for a bit duration  $T$ . By Fourier series expansion, Equation 7.26 can be rewritten as

$$v(t) = \dot{\varphi}_1(t) + \dot{\theta}_1(t) - \frac{d}{dt} \left[ \sum_{m=1}^{\infty} (-1)^m \frac{a^m}{m} \sin \{m\psi(t)\} \right]$$

$W(nT)$ , the output of the integrate-and-dump filter at the sampling time  $nT$  ( $n=0, \pm 1, \pm 2, \dots$ ) or, equivalently, phase shift during  $T$ , becomes

$$W(nT) \approx \varphi_1(nT) - \varphi_1(nT - T) + \frac{d}{dt} \left\{ \sum_{m=1}^{\infty} (-1)^m \frac{2a^m}{m} \cos \left[ \frac{m}{2} \psi_{\sigma}(nT) \right] \sin \left[ \frac{m}{2} \Delta\psi(nT) \right] \right\}$$

where

$$\begin{aligned} \Delta\psi(nT) &= \psi(nT) - \psi(nT - T) \\ \psi_{\sigma}(nT) &= \psi(nT) + \psi(nT - T) \end{aligned}$$

In order for the interference term to disappear,  $\Delta\psi(nT)$  should be

$$\Delta\psi(nT) = \pm 2\pi k, \quad k = 1, 2, 3, \quad (7.29)$$

The output signal is then

$$W(nT) \approx \begin{cases} \varphi_1(nT) - \varphi_1(nT - T) & (R_1(nT) > R_2(nT)) \\ \varphi_2(nT) - \varphi_2(nT - T) & (R_1(nT) < R_2(nT)) \end{cases}$$

This result shows that selection diversity effect is obtained.

A method for getting the condition in Equation 7.29 is phase sweeping or carrier frequency offset [5]. In this method, we set

$$\begin{aligned} \varphi_1(t) &= \frac{\omega_s t}{2} + \frac{\pi\beta}{T} \int_{-\infty}^t s(\tau) d\tau \\ \varphi_2(t) &= \frac{-\omega_s t}{2} + \frac{\pi\beta}{T} \int_{-\infty}^t s(\tau) d\tau \end{aligned}$$

where  $\omega_s$  denotes the offset frequency and  $s(t)$  is a transmit binary NRZ rectangular signal that takes the levels  $+1$  or  $-1$ .  $\beta$  is the modulation index given by  $\beta = 2\Delta f_d T$ ,

where  $\Delta f_d$  denotes the frequency deviation. From Equation 7.29 and the above equations, we get

$$\frac{\omega_s}{2\pi} = \frac{l}{T} \quad (l = 1, 2, 3, \dots)$$

Another method [7] is to use different modulation indices such as

$$\begin{aligned} \varphi_1(t) &= \frac{\pi\beta_1}{T} \int_{-\infty}^{\infty} s(t) dt \\ \varphi_2(t) &= \frac{\pi\beta_2}{T} \int_{-\infty}^{\infty} s(t) dt \end{aligned}$$

From Equation 7.29 and the above equation with  $\int_{-\infty}^{\infty} s(t) dt = T$ , we have

$$\Delta\beta \equiv \beta_2 - \beta_1 = 2l \quad (l = 1, 2, \dots)$$

The other method [6] is to add time-varying signals that meet the conditions of Equation 7.29.

The average received power may be different between the signals from the two base stations. In this case, the argument in Section 7.1.2 cannot be applied. The error rates are analyzed as follows.

Since error rates are not given in closed form for digital FM with limiter-discriminator detection, the following approximate expression is used:

$$P_e(R) \approx \frac{1}{2} \exp \left[ -\alpha \left( \frac{T}{2N_0} \right) R^2 \right]$$

where  $R$  is the received signal envelope,  $N_0$  is the noise power density at the input to the limiter-discriminator, and  $\alpha$  denotes a constant that expresses the appropriate effect of the modulation index and the degradation in the receiver error rate performance. Define  $P_e(R_1, R_2)$  as

$$P_e(R_1, R_2) \approx \begin{cases} \frac{1}{2} \exp \left[ -\alpha \left( \frac{T}{2N_0} \right) R_1^2 \right] & (R_1 > R_2) \\ \frac{1}{2} \exp \left[ -\alpha \left( \frac{T}{2N_0} \right) R_2^2 \right] & (R_1 < R_2) \end{cases} \quad (7.30)$$

The average error rate is given using the above equation and the joint probability density function  $p(R_1, R_2)$  as

$$\begin{aligned} \langle P_e \rangle &= \int_0^{\infty} dR_1 \int_0^{R_1} P_e(R_1, R_2) p(R_1, R_2) dR_2 \\ &\quad + \int_0^{\infty} dR_2 \int_0^{R_2} P_e(R_1, R_2) p(R_1, R_2) dR_1 \end{aligned} \quad (7.31)$$

$p(R_1, R_2)$  is given in [1] as

$$p(R_1, R_2) = \frac{R_1 R_2}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} I_0 \left( \frac{\rho R_1 R_2}{\sigma_1 \sigma_2 (1 - \rho^2)} \right) \exp \left[ -\frac{1}{1 - \rho^2} \left( \frac{R_1^2}{2\sigma_1^2} + \frac{R_2^2}{2\sigma_2^2} \right) \right] \quad (7.32)$$

where  $\sigma_1^2$  and  $\sigma_2^2$  are average powers of the received fading signals. The term  $\rho^2$  (appearing in Section 4.3.3) is a constant that is very near the value of the normalized envelope correlation of the two-branch signals.  $I_0(\cdot)$  is the zeroth-order modified Bessel function of the first kind, which is defined by

$$I_0(z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(-z \cos \theta) d\theta$$

Substituting Equations 7.30 and 7.32 into 7.31 and letting  $R_2/R_1 = t$ , and integrating Equation 7.31 with respect to  $R_1$ ,  $\theta$ , and  $t$ , we have

$$\begin{aligned} \langle P_e \rangle = & \frac{1}{4(1 + \alpha_1 \gamma_1)} \left( 1 - \frac{1 - (\gamma_1/\gamma_2) + \alpha_1 \gamma_1 (1 - \rho^2)}{\{[1 + (\gamma_1/\gamma_2) + \alpha_1 \gamma_1 (1 - \rho^2)]^2 - 4\rho^2 \gamma_1/\gamma_2\}^{1/2}} \right) \\ & + \frac{1}{4(1 + \alpha_2 \gamma_2)} \left( 1 - \frac{1 - (\gamma_2/\gamma_1) + \alpha_2 \gamma_2 (1 - \rho^2)}{\{[1 + (\gamma_2/\gamma_1) + \alpha_2 \gamma_2 (1 - \rho^2)]^2 - 4\rho^2 \gamma_2/\gamma_1\}^{1/2}} \right) \end{aligned}$$

where  $\gamma_1 (\equiv \sigma_1^2 T/N_0)$  and  $\gamma_2 (\equiv \sigma_2^2 T/N_0)$  are the average received signal bit energy to noise power density ratios.

For  $\gamma_1, \gamma_2 \gg 1$ ,

$$\langle P_e \rangle \approx \frac{1}{4\gamma_1 \gamma_2 (1 - \rho^2)} \left( \frac{1}{\alpha_1^2} + \frac{1}{\alpha_2^2} \right)$$

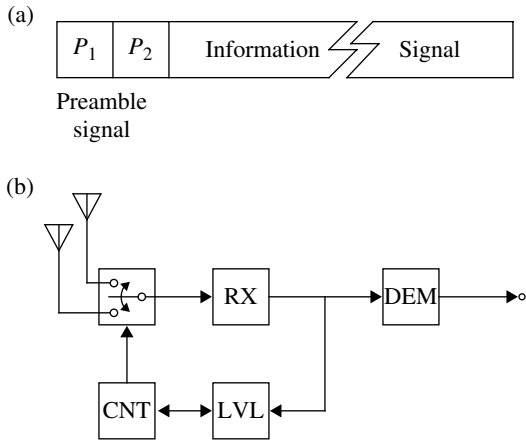
which reduces to Equation 7.11 with  $M=2$  and  $\gamma_1 = \gamma_2 = \gamma_0, \alpha_1 = \alpha_2 = \alpha$ , and  $\rho=0$ .

### 7.1.4 Antenna Selection Diversity System

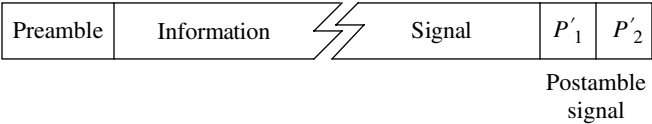
Antenna selection diversity systems have advantages over the others because they require no cophasing circuit for RF or IF combining and no dual receiver for postdetection diversity. However, they generate switching noise, since the received signals at the antenna are not cophased. The switching noise appears as click noise at the demodulator output. In analog systems, the effect of click noise can be mitigated to some degree by use of blanking or sample-and-hold techniques [2]. In digital systems, such techniques are not effective.

A method that avoids the switching noise for digital systems was proposed in [8] and [9], independently. A block diagram of this system is shown in Figure 7.4. The idea for this method originated in noticing that the digital signals are usually sent in





**FIGURE 7.4** Antenna selection diversity receiver. (a) signal format and (b) Antenna selection diversity receiver.



**FIGURE 7.5** Signal format for antenna selection feedback diversity system.

the form of a data block or frame, which consists of a preamble signal and an information signal. The antenna selection is made at the preamble signal period ( $P_1$  and  $P_2$  in Fig. 7.4a). The selection is held during the information signal period. The two antennas are switched to measure each of the received signal levels. The higher signal level antenna is then selected. Although the switching noise appears at the preamble signal period, its effect can be ignored since the preamble signal is a periodic signal for clock or carrier recovery purposes. When the technique is applied to a subscriber (mobile) receiver in a TDMA system, the antenna selection operation can be made at the time slot just before the dedicated time slot. The result is that no switching noise appears in the (preamble) signal.

The proposed method is useful for portable radio terminals, where the equipment needs to be small. The idea of the method was applied in [9] for a transmit diversity system using different frequencies for the uplink and downlink; a diversity antenna is not required for a subscriber radio. In this system, the base station sends postamble signals ( $P'_1$  and  $P'_2$ ) from the antennas in turn, following the information signal, as shown in Figure 7.5. The subscriber receiver examines which part of the postamble signal has the higher level and sends back the results of the examination to the base station. The base station selects the reported antenna to transmit the preamble and information signals in the next time slot. Thus, this method is a type of feedback diversity scheme.

If the base station adopts a diversity system, then both downlink and uplink diversity transmission can be realized. Diversity circuits including antennas may only be required at the base station. When the same frequency is used for the uplink and downlink, then the above technique of transmitting the postamble signal is not necessary. Since the correlation of the uplink and downlink is high, the base station can use the antenna that was selected at the last receiving period for the next transmission. The performance of the proposed system is inferior to that of the other systems since the selected antenna is held for a burst period. The degradation in performance becomes significant in severely fast fading conditions.

After the proposal of this antenna selection diversity system, a timely theoretical analysis on the performance of the system was published [10]. As we discussed in Section 7.1.2, the average error rate can be calculated if the probability density function of the received signal power to noise power density  $\gamma$  is given. Denote the signal amplitude at the output of the selector as  $z$ . The bit energy to noise power ratio  $\gamma$  is given as

$$\gamma = \frac{z^2 T}{N_0} \quad (7.33)$$

where  $T$  is the bit duration and  $N_0$  is the noise power density. Taking the ensemble average of the above equation, we have the average bit energy to noise power density ratio as

$$\langle \gamma \rangle (\equiv \gamma_0) = \frac{\langle z^2 \rangle T}{N_0} \quad (7.34)$$

$\langle z^2 \rangle$  is given as

$$\langle z^2 \rangle = \langle x^2 \rangle + \langle y^2 \rangle = 2R(0) \quad (\langle x^2 \rangle = \langle y^2 \rangle = R(0)) \quad (7.35)$$

where  $x$  and  $y$  are the inphase and quadrature-phase components, respectively, of  $z$ .  $R(\tau)$  is the autocorrelation function of the quadrature components  $x$  and  $y$ , which are Rayleigh faded signals. From Equations 7.33–7.35, we get

$$\gamma = \frac{z^2}{2R(0)} \gamma_0$$

$\gamma_0$  is assumed to take the same value for each branch  $z$  and hence  $\gamma$  are time varying due to fading.

Assume that a selection is made at  $\tau=0$  and let the  $M$  branch amplitudes be  $r_1, r_2, \dots, r_M$ . Let branch  $i$  be selected; then

$$z(\tau) = r_i(\tau),$$

where  $r_i(0) > r_j(0)$  for all  $j(\neq i)$  ( $1, 2, \dots, M$ ) and  $\tau$  is the time elapsed after the selection was made. We assume an identical distribution of the  $M$  branch amplitudes. Assuming

independent fading between the branches, the probability density function of  $z(\tau)$  is given in [10] as

$$f_{z(\tau)} = \sum_{k=0}^{M-1} \binom{M}{k+1} (-1)^k \frac{z(\tau)}{P_k(\tau)} \exp\left[\frac{-z^2(\tau)}{2P_k(\tau)}\right] \quad (7.36)$$

where

$$P_k(\tau) = \frac{R^2(0) - R^2(\tau)k/(k+1)}{R(0)}$$

Departing from [10], we assume (Section 4.3.2)

$$R(\tau) = R_{xx}(\tau) = R_{yy}(\tau) = b_0 J_0(\omega_m \tau)$$

where  $\omega_m$  is the maximum Doppler frequency. Notice that  $f_{z(\tau)}$  is the sum of the Rayleigh distribution functions.

Transforming the variable  $z$  to  $\gamma$  in Equation (7.36), we have

$$f_{\gamma(\tau)} = \sum_{k=0}^{M-1} \binom{M}{k+1} (-1)^k \frac{R(0)}{P_k(\tau)\gamma_0} \exp\left[\frac{-R(0)\gamma(\tau)}{P_k(\tau)\gamma_0}\right] \quad (7.37)$$

Denoting the error rate for a given  $\gamma(\tau)$  as  $P_e[\gamma(\tau)]$ , we have

$$P_e(\tau) = \int_0^\infty P_e[\gamma(\tau)] f_{\gamma(\tau)} d\gamma(\tau) \quad (7.38)$$

The average error rate for our system is then

$$\bar{P} = \frac{1}{T_d} \int_0^{T_d} P_e(\tau) d\tau \quad (7.39)$$

where  $T_d$  is the information time length after a selection is made.

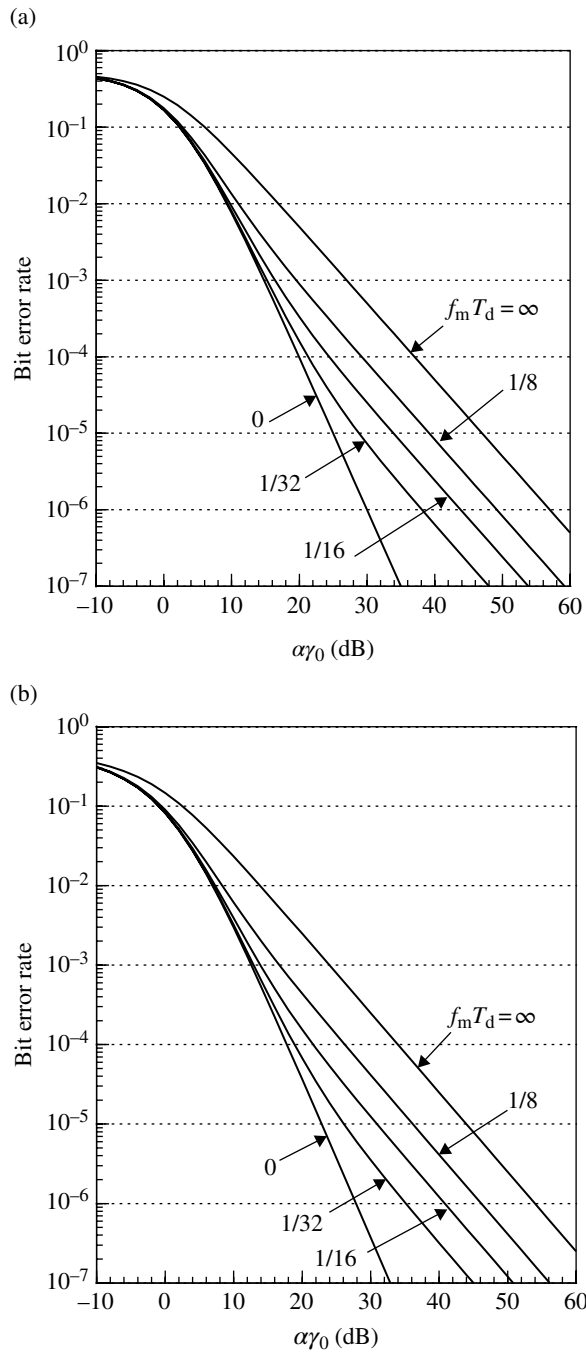
Let us assume that the error rate for a given  $\gamma$  is

$$P_e(\gamma) = \begin{cases} \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) & \text{(coherent detection)} \\ \frac{1}{2} e^{-\alpha\gamma} & \text{(noncoherent detection)} \end{cases} \quad (7.40)$$

$$(7.41)$$

where antenna selection is made at the preamble part of the frame and Rayleigh fading is considered.  $T_d$  is the frame length and  $f_m$  is the maximum Doppler frequency. where  $\alpha$  is a degradation factor. Using Equations 7.37, 7.38, 7.40, and 7.41, we have

$$P_e(\tau) = \sum_{k=0}^{M-1} \binom{M}{k+1} (-1)^k g_k(\tau) \quad (7.42)$$



**FIGURE 7.6** Error rate performance of a 2-branch antenna selection diversity system: (a) coherent detection and (b) noncoherent detection.

where

$$g_k(\tau) = \begin{cases} \frac{1}{2} \left[ 1 - \left( 1 + \frac{R(0)}{P_k(\tau)\alpha\gamma_0} \right)^{-1/2} \right] & (\text{coherent detection}) \end{cases} \quad (7.43)$$

$$g_k(\tau) = \begin{cases} \frac{1}{2} \frac{1}{1 + P_k(\tau)\alpha\gamma_0/R(0)} & (\text{noncoherent detection}) \end{cases} \quad (7.44)$$

If  $R(\tau)=0$ , then  $P_e(\tau)$  yields error rates of a system without diversity. On the other hand, if  $R(\tau)=R(0)$ , then ideal selection diversity is obtained; we can confirm that Equation 7.42 with Equations 7.43 and 7.44 reduces to Equations 7.20 and 7.18, respectively, for the above cases. The error rates calculated with Equations 7.39–7.42 are given in Figure 7.6.

The performance of the antenna selection diversity method was improved in [11] by introducing a selection algorithm based on prediction of received signal levels.

## 7.2 MULTI-INPUT MULTI-OUTPUT SYSTEMS

A multi-input multi-output (MIMO) system has multiple antennas at both transmit and receive sides (Fig. 7.7). This system makes it possible to get space diversity effect as well as a parallel transmission with the space-division multiplexing (SDM) in a radio channel. Here, we discuss the MIMO systems as diversity systems and SDM transmission placing emphasis on explanation of their principles [12].

### 7.2.1 Maximal Ratio Combining Diversity Systems

Space diversity effect is obtained by using multiple antennas at a transmitter and/or a receiver. We discuss the maximal ratio combining diversity that gives the highest signal to noise power ratio (SNR) at the receiver from view point of determining the optimum weights and the attainable SNR for a transmit, a receive, and a transmit and receive diversity system.

**Receive Diversity.** Receive diversity systems called single-input multi-output (SIMO) that use a transmit antenna at the transmitter are described first for simplicity (Fig. 7.8). A combined received signal is expressed as

$$z = \sum_{n=1}^N w_n y_n = \sum_{n=1}^N w_n (h_n x + n_n)$$

where  $h_n$  denotes the propagation coefficients of a channel,  $n_n$  the noise at  $n$ th antenna element, and  $w_n$  the antenna weighting coefficient. We use the complex baseband expression. The signal  $s$  and noise  $n$  at the output of a combining circuit are expressed as

$$s = \sum_{n=1}^N w_n h_n x$$

$$n = \sum_{n=1}^N w_n n_n$$

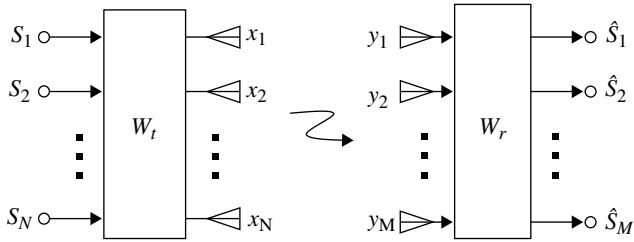


FIGURE 7.7 MIMO system.

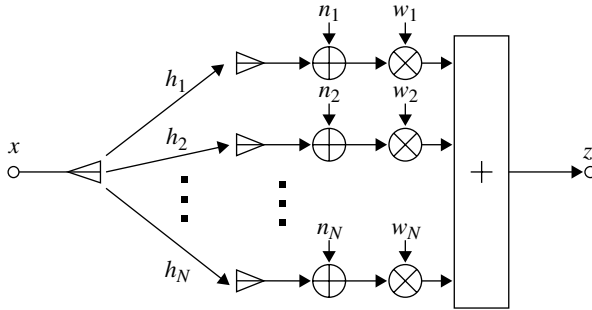


FIGURE 7.8 Receive combining diversity.

Here, we use the column vector notation such as  $\mathbf{h} = (h_1, h_2, \dots, h_N)^T$ , where  $T$  denotes a transpose operation. Then we have

$$\begin{aligned} s &= \mathbf{w}^T \mathbf{h} x \\ n &= \mathbf{w}^T \mathbf{n} \end{aligned} \quad (7.45)$$

Notice that for the same dimensional vectors  $\mathbf{a}$  and  $\mathbf{b}$ ,  $\mathbf{a}^T \mathbf{b} = \mathbf{b}^T \mathbf{a}$  becomes a scalar variable. The average power of the signal and the noise becomes

$$\begin{aligned} S &= \langle |s|^2 \rangle = |\mathbf{w}^T \mathbf{h}|^2 \langle |x|^2 \rangle = |\mathbf{w}^T \mathbf{h}|^2 P_s \\ N &= \langle |\mathbf{w}^T \mathbf{n}|^2 \rangle = \langle \mathbf{w}^T \mathbf{n} (\mathbf{w}^T \mathbf{n})^{*T} \rangle \\ &= \langle \mathbf{w}^T \mathbf{n} \mathbf{n}^{*T} \mathbf{w} \rangle \\ &= \mathbf{w}^T \langle \mathbf{n} \mathbf{n}^H \rangle \mathbf{w}^* \end{aligned} \quad (7.46)$$

where the matrix property  $(\mathbf{ab})^T = \mathbf{b}^T \mathbf{a}^T$  is used. The suffix “ $H$ ” denotes a conjugate transpose and symbol  $\langle \cdot \rangle$  the ensemble average.  $P_s = \langle |x|^2 \rangle$  is the average transmit signal power.

We assume independent noises with zero mean; then we have

$$\langle n_i n_j^* \rangle = \begin{cases} \langle n_i \rangle \langle n_j^* \rangle = 0 & (i \neq j) \\ \langle |n_i|^2 \rangle \equiv N_i & (i = j) \end{cases}$$

Furthermore, we assume the noises have the same average power  $N_i = \sigma^2$ . Thus, we get

$$\langle \mathbf{n} \mathbf{n}^H \rangle = \sigma^2 \mathbf{I}_N \quad (\mathbf{I}_N : \text{unity matrix } (N \times N))$$

Therefore, the noise power given with Equation 7.46 becomes

$$N = \mathbf{w}^T \mathbf{w}^* N_n = \sigma^2 \sum_{n=1}^N |w_n|^2 \quad (7.47)$$

The signal to noise power ratio becomes

$$\frac{S}{N} = \frac{|\mathbf{w}^T \mathbf{h}|^2}{\mathbf{w}^T \mathbf{w}^*} \frac{P_s}{\sigma^2} \quad (7.48)$$

We will get the weights  $\mathbf{w}$  to make  $S/N$  maximum. The Cauchy–Schwarz inequality is known as

$$\left| \sum_{n=1}^N a_n b_n \right|^2 \leq \sum_{n=1}^N |a_n|^2 \cdot \sum_{n=1}^N |b_n|^2$$

The equality holds when  $a_n = k b_n^*$  ( $k$ : constant). We denote the square of the norm as  $\|\mathbf{x}\| \cdot \|\mathbf{x}\| = \|\mathbf{x}\|^2 = \sum_{n=1}^N |x_n|^2 = \mathbf{x}^T \cdot \mathbf{x}^* = \mathbf{x}^H \cdot \mathbf{x}$  for a vector  $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$ . Then the above equation is written with the vector notation as

$$\|\mathbf{a}^T \mathbf{b}\|^2 \leq \|\mathbf{a}\|^2 \|\mathbf{b}\|^2$$

or taking the square root we have

$$\|\mathbf{a}^T \mathbf{b}\| \leq \|\mathbf{a}\| \|\mathbf{b}\|$$

Applying the expression to Equation 7.48, we get

$$\begin{aligned} \frac{S}{N} &\leq \frac{\|\mathbf{w}\|^2 \|\mathbf{h}\|^2}{\|\mathbf{w}\|^2} \frac{P_s}{\sigma^2} = \|\mathbf{h}\|^2 \frac{P_s}{\sigma^2} \\ &= \frac{P_s}{\sigma^2} \sum_{n=1}^N |h_n|^2 \end{aligned} \quad (7.49a)$$

The equality holds when

$$\mathbf{w} = k \mathbf{h}^* \quad (7.49b)$$

Therefore, the optimum weights  $w_n$  ( $n=1,2,\dots,N$ ) become the complex conjugates of the propagation constants  $h_n^*$  excluding a constant  $k$ . Inserting the above equation to Equation 7.45, we have

$$s = k \|\mathbf{h}\|^2 x = k \sum_{n=1}^N |h_n|^2 x$$

We understand the physical meanings of the maximal ratio combining. The signal to noise power ratio  $\gamma_i$  ( $i=1,2,\dots,N$ ) at the output of each antenna element is given as

$$\gamma_i = \frac{\langle |w_i h_i x|^2 \rangle}{|w_i|^2 \sigma^2} = \frac{|h_i|^2}{\sigma^2} \langle |x|^2 \rangle = \frac{P_s}{\sigma^2} |h_i|^2$$

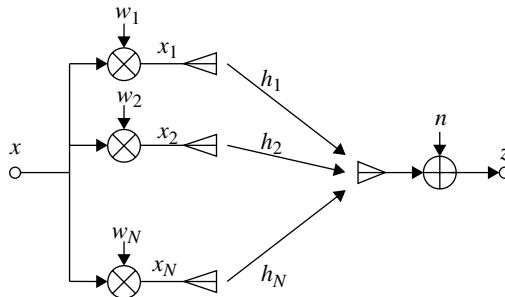
From this relation, the signal to noise power ratio (Equation 7.49a) at the output of the combiner becomes

$$\left. \frac{S}{N} \right|_{\max} = \sum_{i=1}^N \gamma_i$$

The output  $S/N$  becomes the sum of the  $S/N$  at each antenna element.

The above arguments are based on an assumption of equal average noise power. Without the assumption, a similar result is given as follows: we denote the average noise power at each antenna element by  $\sigma_i^2$ . We normalize signals  $s_i$  and noises  $n_i$  at antenna element  $i$  by  $\sigma_i$ . This operation never affects the maximal ratio combining result since the relative relation between signal and noise at an antenna element remains the same. We can apply the above arguments using signal  $s'_i = s_i/\sigma_i$  and noise  $n'_i = n_i/\sigma_i$  since the normalized average power becomes the same as  $\langle |n'_i|^2 \rangle = 1$ .

**Transmit Diversity.** Figure 7.9 shows the system block diagram. We will get the optimum transmit weights that give the maximum signal to noise power ratio at the receiver output. The weights may become infinite without any constraint. A constraint of a fixed total transmit power  $P_t$  is appropriate.  $P_t$  is given as



**FIGURE 7.9** Transmit combining diversity ( $w'_n = w_n/|w|$ ).



$$\begin{aligned}
 P_t &= \left\langle \sum_{n=1}^N |x_n|^2 \right\rangle = \left\langle \sum_{n=1}^N |w_n x|^2 \right\rangle \\
 &= \sum_{n=1}^N |w_n|^2 \langle x^2 \rangle = \|\mathbf{w}\|^2 P_s \quad (P_s = \langle x^2 \rangle)
 \end{aligned}$$

Since  $P_t$  is proportional to  $\|\mathbf{w}\|^2$ , we normalize  $P_t$  with  $\|\mathbf{w}\|^2$  and, therefore, weights  $w_n$  with  $\|\mathbf{w}\|$ . Then the transmit signal vector becomes

$$\mathbf{x} = \frac{\mathbf{w}}{\|\mathbf{w}\|} x$$

The received signal is given as

$$s = \mathbf{h}^T \mathbf{x} = \mathbf{h}^T \frac{\mathbf{w}}{\|\mathbf{w}\|} x$$

We have the received signal average power  $S$  as

$$S = \langle |s|^2 \rangle = \frac{|\mathbf{h}^T \mathbf{w}|^2}{\|\mathbf{w}\|^2} P_s$$

Letting the noise power as  $\sigma^2$ , we have the signal to noise power ratio  $S/N$  as

$$\frac{S}{N} = \frac{|\mathbf{h}^T \mathbf{w}|^2}{\|\mathbf{w}\|^2} \frac{P_s}{\sigma^2}$$

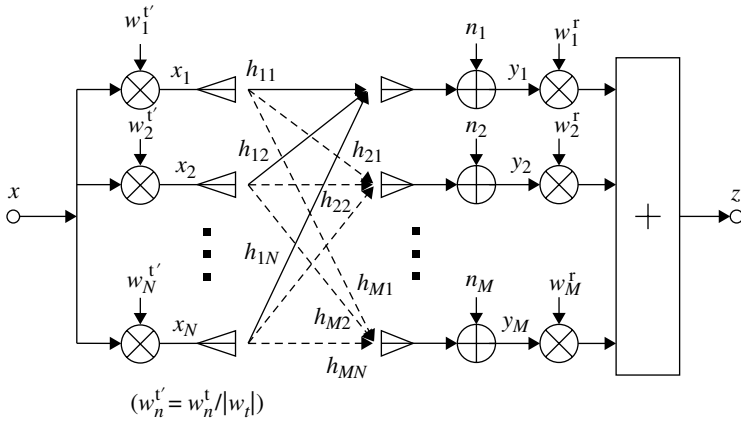
This result is the same as that for the receive diversity (Eq. 7.48,  $\mathbf{w}^T \mathbf{w}^* = \|\mathbf{w}\|^2, \mathbf{w}^T \mathbf{h} = \mathbf{h}^T \mathbf{w}$ ). Therefore, the optimum transmit weights become  $\mathbf{w} = k \mathbf{h}^* / \|\mathbf{w}\|$  ( $k$ : constant). The attainable maximum  $S/N$  is also the same as Equation 7.49a. Thus, the transmit and the receive maximal ratio combining diversity gives the same performance under the same condition. If the absolute values of the propagation constants are the same,  $S/N$  is increased by  $N$  times with  $N$  antennas.

**Transmit and Receive Diversity.** We consider a transmit and receive diversity system with  $N$  transmit- and  $M$  receive-antenna elements (Fig. 7.10). We denote the propagation constant from transmit antenna  $i$  to receive antenna  $j$  as  $h_{ji}$ . The transmit antenna weight  $w_n^t$  is normalized with  $\|\mathbf{w}_t\|$  to keep the transmit power to be constant, where

$$\mathbf{w}_t = (w_1^t, w_2^t, \dots, w_N^t)^T$$

The transmit signal vector becomes

$$\mathbf{x} = \frac{1}{\|\mathbf{w}_t\|} \mathbf{w}_t x$$



**FIGURE 7.10** Transmit and receive combining diversity ( $w_n^{t'} = w_n^t / |w_t|$ ).

The received output signal and noise are given as

$$s = \sum_{m=1}^M w_m^r \left( \sum_{n=1}^N \frac{h_{mn} w_n^t}{\|w_t\|} x \right) = \frac{w_r^T H w_t}{\|w_t\|} x \quad (7.50)$$

$$n = \sum_{m=1}^M w_m^r n_m = w_r^T n$$

where  $H$  ( $M \times N$ ) is the propagation matrix given as

$$H = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22} & \cdots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M1} & h_{M2} & \cdots & h_{MN} \end{pmatrix} \quad (7.51a)$$

and  $w_r = (w_1^r, w_2^r, \dots, w_M^r)^T$ .

The received average signal power becomes

$$S = \langle |s|^2 \rangle = \left\langle \left| \frac{w_r^T H w_t}{\|w_t\|} x \right|^2 \right\rangle$$

$$= \frac{|w_r^T H w_t|^2}{\|w_t\|^2} \langle |x|^2 \rangle \quad (7.51b)$$

$$= \frac{|w_r^T H w_t|^2}{\|w_t\|^2} P_s$$

The noise power is given by changing the notation with Equation 7.47 as

$$N = \mathbf{w}_r^T \mathbf{w}_r^* \sigma^2 = \|\mathbf{w}_r\|^2 \sigma^2 \quad (7.52)$$

From Equations 7.51b and 7.52, we have

$$\frac{S}{N} = \frac{|\mathbf{w}_r^T \mathbf{H} \mathbf{w}_t|^2}{\|\mathbf{w}_t\|^2 \|\mathbf{w}_r\|^2 \sigma^2} \frac{P_s}{\sigma^2} \quad (7.53)$$

We will obtain  $\mathbf{w}_r$  and  $\mathbf{w}_t$  to maximize the above equation. If  $\mathbf{w}_t$  is given first, making  $\mathbf{h} = \mathbf{H} \mathbf{w}_t / \|\mathbf{w}_t\|$  in Equation 7.49b, the optimum  $\mathbf{w}_r$  becomes

$$\mathbf{w}_r = \frac{k' \mathbf{H}^* \mathbf{w}_t^*}{\|\mathbf{w}_t\|} \quad (k' : \text{constant})$$

Inserting this into Equation 7.53, we have

$$\begin{aligned} \frac{S}{N} &= \frac{\|(\mathbf{H}^* \mathbf{w}_t^*)^T \mathbf{H} \mathbf{w}_t\|}{\|\mathbf{w}_t\|^2 \cdot \|\mathbf{H} \mathbf{w}_t\|^2} \frac{P_s}{\sigma^2} \\ &= \frac{\|\mathbf{H} \mathbf{w}_t\|^4}{\|\mathbf{w}_t\|^2 \cdot \|\mathbf{H} \mathbf{w}_t\|^2} \frac{P_s}{\sigma^2} \\ &= \frac{(\mathbf{H}^* \mathbf{w}_t^*)^T \mathbf{H} \mathbf{w}_t}{\|\mathbf{w}_t\|^2} \frac{P_s}{\sigma^2} = \frac{\mathbf{w}_t^H \mathbf{H}^H \mathbf{H} \mathbf{w}_t}{\|\mathbf{w}_t\|^2} \frac{P_s}{\sigma^2} \end{aligned}$$

Applying the Cauchy–Schwarz inequality, we get

$$\frac{S}{N} \leq \frac{\|\mathbf{w}_t\| \cdot \|\mathbf{H}^H \mathbf{H} \mathbf{w}_t\|}{\|\mathbf{w}_t\|^2} \frac{P_s}{\sigma^2} = \frac{\|\mathbf{H}^H \mathbf{H} \mathbf{w}_t\|}{\|\mathbf{w}_t\|} \frac{P_s}{\sigma^2} \quad (7.54)$$

The equality holds with the following equation:

$$\mathbf{w}_t = k \mathbf{H}^H \mathbf{H} \mathbf{w}_t$$

This relation is rewritten as

$$\mathbf{H}^H \mathbf{H} \mathbf{w}_t = \lambda \mathbf{w}_t \quad (\lambda = 1/k) \quad (7.55)$$

where  $\mathbf{w}_t$  is the eigenvector of the matrix  $\mathbf{H}^H \mathbf{H}$  and  $\lambda$  is its eigenvalue. Multiplying  $\mathbf{w}_t^H$  from the left side in the above equation, we have

$$\mathbf{w}_t^H \mathbf{H}^H \mathbf{H} \mathbf{w}_t = \|\mathbf{H} \mathbf{w}_t\|^2 = \lambda \|\mathbf{w}_t\|^2 \geq 0$$

Thus, the eigenvalue  $\lambda$  takes a positive real number.

Applying Equations 7.55 to 7.54, we get the maximum  $S/N$  as

$$\left. \frac{S}{N} \right|_{\max} = \lambda \frac{P_s}{\sigma^2} \quad (7.56)$$

If the eigenvector  $\mathbf{w}_t$  that has the maximum eigenvalue  $\lambda$  is used, the signal to noise power  $S/N$  at the receiver output becomes maximum. Therefore, this system attains the maximum transmission capacity for given transmit signal power and receive noise power.

Applying the receive optimum weights  $\mathbf{w}_r = k' \mathbf{H}^* \mathbf{w}_t^* / \|\mathbf{w}_t\|$  with  $k' = 1$  to Equation 7.50, the receive output signal becomes

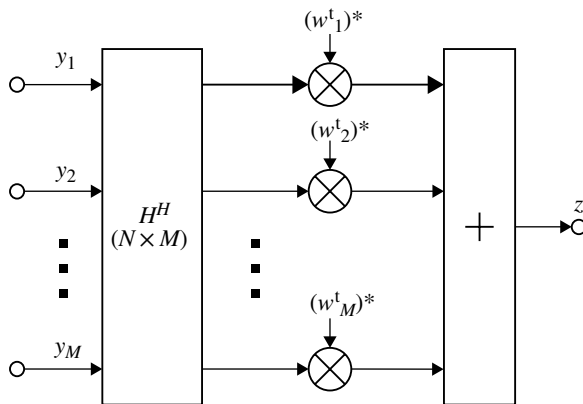
$$\begin{aligned} s &= \mathbf{H}^* \frac{\mathbf{w}_t^*}{\|\mathbf{w}_t\|} \frac{\mathbf{H} \mathbf{w}_t}{\|\mathbf{w}_t\|} x \\ &= \frac{\mathbf{w}_t^H}{\|\mathbf{w}_t\|} \frac{\mathbf{H}^H \mathbf{H} \mathbf{w}_t}{\|\mathbf{w}_t\|} x \end{aligned} \quad (7.57a)$$

This equation is rewritten with Equation 7.55 as

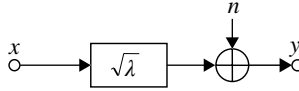
$$\begin{aligned} s &= \frac{\mathbf{w}_t^H \lambda \mathbf{w}_t}{\|\mathbf{w}_t\|^2} x \\ &= \lambda x \end{aligned} \quad (7.57b)$$

Thus, the receive output signal average power becomes  $S = \langle |s|^2 \rangle = \lambda^2 P_s$  and the average noise power becomes  $N = \|\mathbf{w}_r\|^2 \sigma^2 = \lambda \sigma^2$  ( $k' = 1$ ).

Equation 7.57a gives a receiver for the optimum combining transmit and receive diversity system as shown in Figure 7.11. And Equation 7.57b gives its equivalent circuit as shown in Figure 7.12.



**FIGURE 7.11** Receive circuit for optimum combining transmit and receive diversity.



**FIGURE 7.12** Equivalent circuit for optimum combining transmit and receive diversity.

The above arguments are developed under the assumption that transmit weights  $\mathbf{w}_t$  are given first. A similar treatment is possible when the receive weights  $\mathbf{w}_r$  are given first. The result is given as follows. The optimum weights  $\mathbf{w}_r$  are given as the eigenvector of a matrix  $\mathbf{H}^* \mathbf{H}^T$  as

$$\mathbf{H}^* \mathbf{H}^T \mathbf{w}_r = \lambda \mathbf{w}_r$$

Taking the complex conjugate of the above equation, we have

$$\mathbf{H} \mathbf{H}^H \mathbf{w}_r^* = \lambda \mathbf{w}_r^* \quad (\lambda : \text{real})$$

Thus, the optimum  $\mathbf{w}_r$  is the complex conjugate of the eigenvector of  $\mathbf{H} \mathbf{H}^H$ . The optimum transmit weights become

$$\mathbf{w}_t = k \mathbf{H}^H \mathbf{w}_r^*$$

The maximum  $S/N$  at the output of the receiver is the same as the former result (Eq. 7.56).

The transmit signal vector becomes

$$\mathbf{x} = \frac{\mathbf{H}^H \mathbf{w}_r^*}{\|\mathbf{H}^H \mathbf{w}_r^*\|} x \quad (7.58)$$

where normalization is carried out to keep the transmit power fixed. Normalizing the receive weights  $\mathbf{w}_r$ , we have the received signal as

$$s = \frac{\mathbf{w}_r^T}{\|\mathbf{w}_r\|} \mathbf{H} \mathbf{x} \quad (7.59a)$$

$$= \frac{\mathbf{w}_r^T}{\|\mathbf{w}_r\|} \frac{\mathbf{H} \mathbf{H}^H \mathbf{w}_r^*}{\|\mathbf{H}^H \mathbf{w}_r^*\|} x = \sqrt{\lambda} x \quad (7.59b)$$

The receive output signal power becomes  $S = \langle |s|^2 \rangle = \lambda P_s \left( P_s = \langle |x|^2 \rangle \right)$ . The noise power is  $N = \sigma^2$ , since  $\mathbf{w}_r$  is normalized.

From Equations 7.58 and 7.59a, we have a transmit and receive diversity system, where transform  $\mathbf{H}^H$  is done at the transmitter.

In the above arguments, we need a linear transform with the matrix  $\mathbf{H}^H$  in addition to multiplications with the weights  $\mathbf{w}_t$  and  $\mathbf{w}_r$ . If we use the following method called singular value decomposition [13], the linear transform can be omitted. A matrix  $\mathbf{H}$  ( $n \times m$ ) with rank  $r$  can be expanded as

$$\begin{aligned}\mathbf{H} &= \mu_1 \mathbf{v}_1 \mathbf{u}_1^H + \mu_2 \mathbf{v}_2 \mathbf{u}_2^H + \cdots + \mu_r \mathbf{v}_r \mathbf{u}_r^H \\ &= \mathbf{V} \Delta^H \mathbf{U}^H\end{aligned}\quad (7.60)$$

where  $\mu_j = \sqrt{\lambda_j}$  ( $j=1,2,\dots,r$ ),  $\lambda_j$  is a nonzero eigenvalue of a matrix  $\mathbf{H}^H \mathbf{H}$ , and

$$\Delta = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_r \end{pmatrix}$$

$$\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r),$$

$$\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r)$$

$$\mathbf{u}_j = (u_{j1}, u_{j2}, \dots, u_{jm})^T,$$

$$\mathbf{v}_j = (v_{j1}, v_{j2}, \dots, v_{jn})^T$$

$$\mathbf{U}^H \mathbf{U} = \mathbf{V}^H \mathbf{V} = \mathbf{I}_r$$

Then we have

$$\mathbf{u}_i^H \mathbf{u}_j = \mathbf{v}_i^H \mathbf{v}_j = \begin{cases} 1 & (i=j) \\ 0 & (i \neq j) \end{cases}$$

Vectors  $\mathbf{u}_i$  and  $\mathbf{u}_j$  that satisfy Equation 7.60 meet the following relations:

$$\mathbf{H} \mathbf{u}_j = \mu_j \mathbf{v}_j \quad \text{and} \quad \mathbf{H}^H \mathbf{v}_j = \mu_j \mathbf{u}_j \quad (7.61a)$$

$$\mathbf{H}^H \mathbf{H} \mathbf{u}_j = \lambda_j \mathbf{u}_j \quad \text{and} \quad \mathbf{H} \mathbf{u}_j = \mu_j \mathbf{v}_j \quad (7.61b)$$

$$\mathbf{H} \mathbf{H}^H \mathbf{v}_j = \lambda_j \mathbf{v}_j \quad \text{and} \quad \mathbf{H}^H \mathbf{v}_j = \mu_j \mathbf{u}_j \quad (7.61c)$$

From Equation 7.61a, we have

$$(\mathbf{H} \mathbf{u}_j)^H = \mathbf{u}_j^H \mathbf{H}^H = \mu_j \mathbf{v}_j^H$$

If we let  $\mathbf{u}_j = \frac{\mathbf{w}_t}{\|\mathbf{w}_t\|}$ , Equation 7.57a is written as

$$s = \frac{\mathbf{w}_t^H}{\|\mathbf{w}_t\|} \mathbf{H}^H \mathbf{y} = \mu_j \mathbf{v}_j^H \mathbf{y} \quad \left( \mathbf{y} = \frac{\mathbf{H} \mathbf{W}_t}{\|\mathbf{W}_t\|} \mathbf{x} \right)$$

Thus, we understand that the multiplication with weights  $\mathbf{v}_j^H$  is sufficient at the receiver ( $\mu_j$  affect only the receiver gain).

As is seen from Equations 7.61b and 7.61c, matrices  $\mathbf{H}^H \mathbf{H}$  and  $\mathbf{H} \mathbf{H}^H$  have the same eigenvalues. Therefore, the same maximum value of  $S/N$  is given with the multiplications by  $\mathbf{H}^H$  at either the transmitter or the receiver.

### 7.2.2 Space–Time Codes

A transmit diversity system requires knowledge of propagation characteristics at the transmitter side. The space–time codes are known as a method without necessity of the knowledge. In a transmit and receive diversity system, transmit and receive signals at antenna elements are just multiplied with weights. In a space–time code, signals are coded in time domain in addition to the weight multiplications. As coding in time domain, two methods are known: space–time block code (STBC) [14], which uses a block coding, and space–time trellis code (STTC), which uses a convolutional code with the maximal likelihood decoding [15]. Here, we explain only the STBC for simplicity.

We consider a MIMO system with two transmit antennas and  $N$  receive antennas. Transmit signal is denoted as  $s[m]$  at symbol times  $m$ . The transmit signals from each antenna are coded as.

$$\begin{aligned}x_1[m] &= s[m] \\x_2[m] &= s[m+1] \\x_1[m+1] &= -s^*[m+1] \\x_2[m+1] &= s^*[m]\end{aligned}$$

where the suffixes 1 and 2 express each antenna element. The same signal is transmitted from each antenna at different time. The received signal at  $i$ th receive antenna element is expressed as

$$\begin{aligned}y_i[m] &= h_{i1}x_1[m] + h_{i2}x_2[m] + n_i[m] \\&= h_{i1}s[m] + h_{i2}s[m+1] + n_i[m] \\y_i[m+1] &= h_{i1}x_1[m+1] + h_{i2}x_2[m+1] + n_i[m+1] \\&= -h_{i1}s^*[m+1] + h_{i2}s^*[m] + n_i[m+1]\end{aligned}$$

where  $h_{i1}$  and  $h_{i2}$  are propagation constants from the transmit antenna 1 and 2 to a receive antenna  $i$ , respectively, and  $n_i[m]$  is the noise at the receiver antenna  $i$ . We assume the propagation constants are known at the receiver. At the receiver, the following operation is done:

$$\begin{aligned}h_{i1}^*y_i[m] + h_{i2}^*y_i[m+1] &= (|h_{i1}|^2 + |h_{i2}|^2)s[m] + h_{i1}^*n_i[m] + h_{i2}^*n_i[m+1] \\h_{i2}^*y_i[m] - h_{i1}^*y_i[m+1] &= (|h_{i1}|^2 + |h_{i2}|^2)s[m+1] + h_{i2}^*n_i[m] + h_{i1}^*n_i[m+1]\end{aligned}$$

We see that signals  $s[m]$  and  $s[m+1]$  are separately obtained. The signal to noise power ratio  $\gamma_i$  at antenna  $i$  becomes

$$\begin{aligned}\gamma_i &= \frac{(|h_{i1}|^2 + |h_{i2}|^2)^2 \langle |s[m]|^2 \rangle}{(|h_{i1}|^2 + |h_{i2}|^2) \sigma_i^2} \\&= \frac{(|h_{i1}|^2 + |h_{i2}|^2) \langle |s|^2 \rangle}{\sigma_i^2}\end{aligned}$$

where we assumed that the signal power takes a constant level  $\langle |s|^2 \rangle$  irrespective of each data symbol and noises (with zero mean) at different symbol time are independent with each other. Comparing the above result with that of Equation 7.49a, we see two-branch MRC diversity effect is obtained. The transmit signal power  $P_s$  becomes two times of  $\langle |s|^2 \rangle$  since the same signal is sent twice. Consequently, we have

$$\gamma_i = \left( |h_{i1}|^2 + |h_{i2}|^2 \right) \frac{P_s}{2\sigma_i^2}$$

With the same transmit power, the obtained  $S/N$  becomes one-half of that for an MRC transmit diversity. This is due to the fact that the propagation constant is unknown at the transmitter.

With maximal ratio combining of the outputs of each antenna element, a  $2N$  branch diversity effect on  $S/N$  (Section 7.2.1) is given as

$$\begin{aligned} \frac{S}{N} &= \sum_{i=1}^N \left( |h_{i1}|^2 + |h_{i2}|^2 \right) \frac{P_s}{\sigma_i^2} \\ &= \sum_{i=1}^N \gamma_i \end{aligned}$$

We should notice the transmission speed is not sacrificed, since the two signals  $s[m]$  and  $s[m+1]$  are sent at times  $m$  and  $m+1$ . This method is applicable for a system with more than two transmit antennas. However, the transmission speed is sacrificed to keep orthogonality between signals [16].

### 7.2.3 SDM in MIMO Systems

For simplicity, we consider a MIMO system with two antennas both at transmit and receive sides (Figure 7.13). Noise is not considered at the time being. The independent two transmit signal sequences  $s_1$  and  $s_2$  are transmitted as signal sequences  $x_1$  and  $x_2$  from transmit antennas #1 and #2, respectively, after passing through a linear combining circuit.

The transmit signals are multiplied with propagation constants  $h_{ij}$  ( $i, j = 1, 2$ ) (complex notation) and received at antennas #1 and #2. The received signals are applied to a linear combining circuit and becomes output signals  $\hat{s}_1$  and  $\hat{s}_2$ . In an ideal

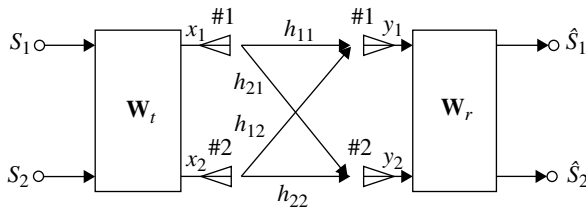


FIGURE 7.13 MIMO system with two antennas for transmitter and receiver.



condition, we get  $\hat{s}_1 = s_1$  and  $\hat{s}_2 = s_2$ . Actually, however, the signals  $\hat{s}_1$  and  $\hat{s}_2$  are degraded due to noise and interference between the signals.

We will explain using mathematical expressions in the following (Fig. 7.13).

The received signals  $y_1$  and  $y_2$  at antennas 1 and 2, respectively, are expressed as

$$y_1 = h_{11}x_1 + h_{12}x_2$$

$$y_2 = h_{21}x_1 + h_{22}x_2$$

Using matrix notation, we have

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

In a linear combining circuit (Fig. 7.14), we have similar expressions as

$$x'_1 = w_{11}x_1 + w_{12}x_2$$

$$x'_2 = w_{21}x_1 + w_{22}x_2$$

$$\begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

For the first, we consider an ideal case such as  $h_{11} = h_{22} = 1$  and  $h_{12} = h_{21} = 0$ . Then we have  $y_1 = x_1$  and  $y_2 = x_2$  without interference between the signals. Letting  $x_1 = s_1$ ,  $x_2 = s_2$ ,  $\hat{s}_1 = y_1$  and  $\hat{s}_2 = y_2$ , and therefore  $w_{11}^t = w_{22}^t = w_{11}^r = w_{22}^r = 1$ , and  $w_{12}^t = w_{21}^t = w_{12}^r = w_{21}^r = 0$ , we get  $\hat{s}_1 = s_1$  and  $\hat{s}_2 = s_2$ .

Here, the upper subscripts t and r of the weights denote weights at the transmit and the receive sides, respectively. The above case means, for example, a parallel independent transmission system for signals  $s_1$  and  $s_2$  with two ideal coaxial cables (without loss and coupling between the cables). In a wireless transmission system, if parabola antennas with a sharp directivity pattern are used, a closely ideal space division multiplexed becomes possible similarly. In a mobile radio communication system, however, interference between signals is inevitable ( $h_{12} \neq 0, h_{21} \neq 0$ ), since

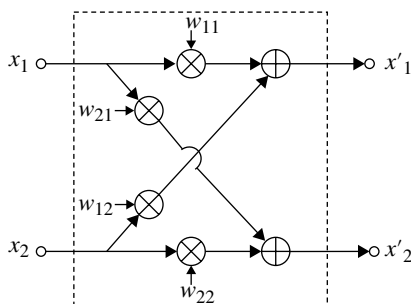


FIGURE 7.14 Linear combination circuit ( $2 \times 2$ ).

omnidirectional antennas are usually used. The MIMO SDM is a technique that makes parallel transmission possible by suppressing the interference with signal processing.

**Inverse Matrix Method.** The first method is to implement a receiver linear combining circuit that has the inverse matrix of the propagation matrix, that is,  $\mathbf{w}^r = \mathbf{H}^{-1}$ , where  $\mathbf{H}$  is the propagation matrix and  $\mathbf{w}^r$  is the receive weighting matrix. We have

$$\begin{aligned} \begin{pmatrix} \hat{s}_1 \\ \hat{s}_2 \end{pmatrix} &= \begin{pmatrix} w_{11}^r & w_{12}^r \\ w_{21}^r & w_{22}^r \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} w_{11}^r & w_{12}^r \\ w_{21}^r & w_{22}^r \end{pmatrix} \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \mathbf{H}^{-1} \mathbf{H} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \end{aligned}$$

Thus, we get  $\hat{s}_1 = s_1$  and  $\hat{s}_2 = s_2$  by letting  $x_1 = s_1$  and  $x_2 = s_2$ . The inverse matrix operation is possible at the transmitter side, since the transmission system as a whole is a linear system.

The above argument assumed tacitly the existence of the inverse matrix for  $\mathbf{H}$ . We need the condition  $\det \mathbf{H} = h_{11}h_{22} - h_{12}h_{21} \neq 0$  for validness of the assumption. Let us consider a case where the condition is not met, that is, the signal separation becomes impossible.

Since this is a case with  $h_{11}h_{22} = h_{12}h_{21}$ , we have  $h_{11} = \alpha h_{12}$  and  $h_{22} = \alpha^{-1}h_{21}$  (or  $h_{11} = \alpha h_{21}$  and  $h_{22} = \alpha^{-1}h_{12}$ ) for an example. Then we have

$$\begin{aligned} y_1 &= \alpha h_{12}x_1 + h_{12}x_2 = h_{12}(\alpha x_1 + x_2) \\ y_2 &= h_{21}x_1 + \alpha^{-1}h_{21}x_2 = h_{21}\alpha^{-1}(\alpha x_1 + x_2) \end{aligned}$$

From this result, we get

$$y_2 = \frac{\alpha^{-1}h_{21}}{h_{12}} y_1$$

We see that  $y_2$  and  $y_1$  are not independent with each other and the signal separation becomes impossible. A physical meaning of this case is easy to understand especially for  $\alpha = 1$ .

We consider an actual example where signal separation becomes difficult. If the distance between transmit and receive antennas is much longer than antenna element spacing and a receive radio wave becomes a plane wave in a line-of-sight channel, the propagation constant is expressed as

$$h_{ij} \cong h_0 e^{j \frac{2\pi}{\lambda} l_{ij}} \quad (i, j = 1, 2) \quad (h_0 : \text{constant}) \quad (7.62)$$

where  $l_{ij} = y_i - x_j$  is a distance projected on a line from transmit antenna  $i$  to receive antenna  $j$  (Fig. 7.15). This result means that the effect of the distance perpendicular to the line on the plane wave propagation constant diminishes as distance  $d$  becomes infinite. Thus,  $l_{11} + l_{22} \cong l_{12} + l_{21}$  is given, and from Equation 7.62,  $h_{11}h_{22} \cong h_{12}h_{21}$  holds.

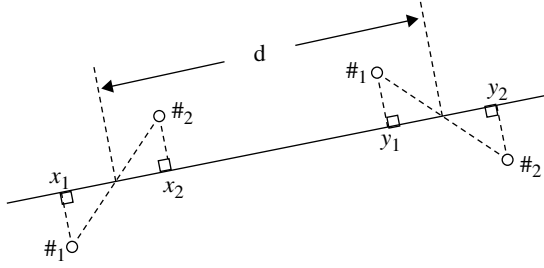


FIGURE 7.15 MIMO antenna arrangement.

**Eigenvector Method.** The second method to separate interfering signals is diagonalization of propagation matrix  $\mathbf{H}$ . The output vector  $[y_{i1}^e \ y_{i2}^e]^T$  for an input eigenvector  $[x_{i1}^e \ x_{i2}^e]^T$  ( $i = 1, 2$ ; T: transpose) is expressed as

$$\begin{bmatrix} y_{i1}^e \\ y_{i2}^e \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_{i1}^e \\ x_{i2}^e \end{bmatrix} = \lambda_i \begin{bmatrix} x_{i1}^e \\ x_{i2}^e \end{bmatrix} \quad (i = 1, 2)$$

where  $\lambda_i$  is an eigenvalue corresponding to the eigenvector. We transform the transmit signals  $s_1$  and  $s_2$  as

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = s_1 \begin{pmatrix} x_{11}^e \\ x_{12}^e \end{pmatrix} + s_2 \begin{pmatrix} x_{21}^e \\ x_{22}^e \end{pmatrix} = \begin{bmatrix} x_{11}^e & x_{21}^e \\ x_{12}^e & x_{22}^e \end{bmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

Then transformed signals  $x_1$  and  $x_2$  are transmitted from antennas 1 and 2, respectively. Then the received signals are given as

$$\begin{aligned} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} &= \lambda_1 s_1 \begin{pmatrix} x_{11}^e \\ x_{12}^e \end{pmatrix} + \lambda_2 s_2 \begin{pmatrix} x_{21}^e \\ x_{22}^e \end{pmatrix} \\ &= \begin{bmatrix} x_{11}^e & x_{21}^e \\ x_{12}^e & x_{22}^e \end{bmatrix} \begin{bmatrix} \lambda_1 s_1 \\ \lambda_2 s_2 \end{bmatrix} \\ &= \begin{bmatrix} x_{11}^e & x_{21}^e \\ x_{12}^e & x_{22}^e \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \end{aligned}$$

Multiplying a matrix  $\begin{bmatrix} x_{11}^e & x_{21}^e \\ x_{12}^e & x_{22}^e \end{bmatrix}^{-1}$  from the left side of the above equation, we have

$$\begin{aligned} \begin{pmatrix} \hat{s}_1 \\ \hat{s}_2 \end{pmatrix} &\equiv \begin{bmatrix} x_{11}^e & x_{21}^e \\ x_{12}^e & x_{22}^e \end{bmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 s_1 \\ \lambda_2 s_2 \end{bmatrix} \end{aligned}$$

Thus, we get  $\hat{s}_i = \lambda_i s_i$  that is the transmit signal  $s_i$  multiplied with the eigenvalue  $\lambda_i$ . The system block diagram is shown in Figure 7.13 with

$$\mathbf{w}^t = \begin{bmatrix} x_{11}^e & x_{21}^e \\ x_{12}^e & x_{22}^e \end{bmatrix}, \quad \mathbf{w}^r = \begin{bmatrix} x_{11}^e & x_{21}^e \\ x_{12}^e & x_{22}^e \end{bmatrix}^{-1}$$

The system operation as a whole is expressed as

$$\begin{aligned} \begin{pmatrix} \hat{s}_1 \\ \hat{s}_2 \end{pmatrix} &= \begin{bmatrix} x_{11}^e & x_{21}^e \\ x_{12}^e & x_{22}^e \end{bmatrix}^{-1} \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_{11}^e & x_{21}^e \\ x_{12}^e & x_{22}^e \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \end{aligned}$$

Thus, we see that this method is to diagonalize the propagation matrix with the eigenvector. The eigenvalues are given as

$$\lambda_i = \frac{h_{11} + h_{22} \pm \sqrt{(h_{11} + h_{22})^2 - 4(h_{11}h_{22} - h_{12}h_{21})}}{2} \quad (\pm \text{ correspond to } i = 1, 2)$$

If  $h_{11}h_{22} = h_{12}h_{21}$ , then  $\lambda_1 = h_{11} + h_{22}$ ,  $\lambda_2 = 0$  and the SDM is impossible. If  $h_{12} = 0$  or  $h_{21} = 0$ , we have  $\lambda_1 = h_{11}$ ,  $\lambda_2 = h_{22}$ .

This method as well as the inverse matrix method is applicable to a system with more than 2 antenna elements as long as the transmitter and the receiver use antenna element of the same number. The maximum number of multiplexing signals is the rank of the propagation matrix.

**Eigen-Beam Method.** This method (E-SDM) is applicable to a system that uses different number antennas at the transmit and the receive sides. Furthermore, this method gives the maximum transmission capacity. We assume  $N$  and  $M$  antenna elements at the transmitter and the receiver, respectively. A propagation matrix  $\mathbf{H}(M \times N)$  is given with Equation 7.51a. This method uses eigenvectors of a matrix  $\mathbf{H}^H \mathbf{H}$  (or  $\mathbf{H} \mathbf{H}^H$ ) to transmit and receive signal in parallel. The eigenvector and the eigenvalues of  $\mathbf{H}^H \mathbf{H}$  appeared in the discussion of a transmit and receive MRC diversity system (Section 7.2.1). This system provides SDM while keeping the maximum signal to noise power ratio  $S/N$ . The fact that the SDM is possible is understood, since this system is the same as the eigenvector method in principle. We explain the system using mathematical expressions.

The eigenvalues of the matrix  $\mathbf{H}^H \mathbf{H}$  ( $N \times N$ ) are nonnegative real numbers (Section 7.2.1). We denote these in a descending order as

$\lambda_1 > \lambda_2 > \dots > \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_N = 0$ , where  $r (\leq \min(N, M))$  is the rank of the matrix  $\mathbf{H}$ . It is known that  $\text{rank}(\mathbf{H}) = \text{rank}(\mathbf{H}^H \mathbf{H})$ . We denote an eigenvector  $\mathbf{u}_i$  and its corresponding eigenvalue as  $\lambda_i$ . Then we have

$$\mathbf{H}^H \mathbf{H} \mathbf{u}_i = \lambda_i \mathbf{u}_i \quad (i = 1, 2, \dots, N) \quad (7.63)$$

Since the eigenvector has an arbitrary multiplying constant, we normalize the norm of  $\mathbf{u}_i$  as  $\|\mathbf{u}_i\|=1$ . A data sequence that uses  $\mathbf{u}_i$  is denoted as  $s_i$  ( $i=1,2, \dots,k$ ), where  $k(\leq r)$  is the total number of the space-division multiplexed signals. A signal vector  $\mathbf{x}$  transmitted from antennas is expressed as

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} = s_1 \begin{pmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1N} \end{pmatrix} + s_2 \begin{pmatrix} u_{21} \\ u_{22} \\ \vdots \\ u_{2N} \end{pmatrix} + \cdots + s_k \begin{pmatrix} u_{k1} \\ u_{k2} \\ \vdots \\ u_{kN} \end{pmatrix}$$

$$= \begin{pmatrix} u_{11} & u_{21} & \cdots & u_{k1} \\ u_{12} & u_{22} & \cdots & u_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{1N} & u_{2N} & \cdots & u_{kN} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_k \end{pmatrix}$$

We use a vector notation as

$$\mathbf{x} = \mathbf{U}\mathbf{s}, \text{ where } \mathbf{U} = \begin{pmatrix} u_{11} & u_{21} & \cdots & u_{k1} \\ u_{12} & u_{22} & \cdots & u_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{1N} & u_{2N} & \cdots & u_{kN} \end{pmatrix}$$

The signal vector  $\mathbf{y}$  at receiving antennas is expressed as

$$\begin{aligned} \mathbf{y} &= \mathbf{H}\mathbf{x} \\ &= \mathbf{H}\mathbf{U}\mathbf{s} \end{aligned}$$

The operation of the receiver is described as

$$\hat{\mathbf{s}} = \mathbf{U}^H \mathbf{H}^H \mathbf{y} \quad (7.64a)$$

$$= \mathbf{U}^H \mathbf{H}^H \mathbf{H} \mathbf{U} \mathbf{s} \quad (7.64b)$$

Considering Equation 7.63, we have

$$\begin{aligned} \mathbf{H}^H \mathbf{y} &= \mathbf{H}^H \mathbf{H} \mathbf{U} \mathbf{s} \\ &= \lambda_1 s_1 \mathbf{u}_1 + \lambda_2 s_2 \mathbf{u}_2 + \cdots + \lambda_k s_k \mathbf{u}_k \\ &= \mathbf{U} \begin{bmatrix} \lambda_1 s_1 \\ \lambda_2 s_2 \\ \vdots \\ \lambda_k s_k \end{bmatrix} \end{aligned}$$

Inserting this into Equation 7.64a, we get

$$\hat{\mathbf{s}} = \mathbf{U}^H \mathbf{U} \begin{bmatrix} \lambda_1 s_1 \\ \lambda_2 s_2 \\ \vdots \\ \lambda_k s_k \end{bmatrix}$$

Since  $\mathbf{H}^H \mathbf{H}$  is a normal matrix, any two different eigenvectors are orthogonal with each other, that is,

$$\mathbf{u}_i^H \mathbf{u}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

where it is reflected that  $\mathbf{u}_i$  are normalized for their norm to be 1.

Therefore,  $\mathbf{U}^H \mathbf{U} = \mathbf{I}_k$ , where  $\mathbf{I}_k$  ( $k \times k$ ) is a unit matrix. Then we have

$$\hat{\mathbf{s}} = \begin{bmatrix} \lambda_1 s_1 \\ \lambda_2 s_2 \\ \vdots \\ \lambda_k s_k \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_k \end{bmatrix}$$

The data are separated at the receiver output. We can apply the other method that uses the singular value decomposition as discussed in Section 7.2.1.

If we put  $k=N$  in the above arguments, the matrix  $\mathbf{U}$  ( $N \times N$ ) becomes a unitary matrix. Since  $\mathbf{U}^H \mathbf{U} = \mathbf{I}_N$  for a unitary matrix  $\mathbf{U}$  ( $N \times N$ ), we have  $\mathbf{U}^H = \mathbf{U}^{-1}$ . Inserting this to Equation 7.64b, we have

$$\hat{\mathbf{s}} = \mathbf{U}^{-1} \mathbf{H}^H \mathbf{H} \mathbf{U} \mathbf{s}$$

Therefore, this method falls into a category with diagonalization as the same as the eigenvector method. The difference between them is that matrix  $\mathbf{H}^H \mathbf{H}$  is used instead of  $\mathbf{H}$ . The MRC effect is obtained through this fact.

We considered so far the matrix  $\mathbf{H}^H$  is multiplied at a transmitter. We obtain a similar space-division multiplexed MIMO system where the multiplication with  $\mathbf{H}^H$  is moved to the receiver.

Signal to noise power ratio ( $S/N$ ) is given by Equation 7.56 obtained for an MRC system by placing signal power of each signal sequence as  $P_s$ .

The output  $S/N$  for different signal sequences becomes different depending on the eigenvector and hence the eigenvalue assigned to the signal sequence.

**Transmit Power Allocation.** Allocating transmit power of a given total power to each data sequence is another task in a MIMO system. The results depend on criteria selected for system performance. Transmission capacity and average error rates are taken for the criteria. We consider first power allocation that yields the maximum transmission capacity. The Shannon capacity is given as

$$C = W \log(\gamma + 1) \quad (\text{bits/s})$$

where  $W$  denotes bandwidth,  $\gamma$  is the signal to noise power ratio, and  $\log_2$  is understood for  $\log$ .

The capacity normalized with  $W$  becomes

$$C' = \frac{C}{W} = \log(\gamma + 1) \quad (\text{bits/s/Hz})$$

Denoting transmit power by  $p_i$  ( $i=1,2, \dots, K$ ) for data sequences  $i$  and using Equation 7.56, we have

$$c'_i = \log\left(\frac{p_i}{\sigma^2} \lambda_i + 1\right)$$

The total transmission capacity becomes

$$J = \sum_{i=1}^K c'_i = \sum_{i=1}^K \log\left(\frac{p_i}{\sigma^2} \lambda_i + 1\right) \quad (7.65)$$

We determine  $p_i$  ( $> 0$ ) that makes  $J$  maximum under a constraint of  $P_x = \sum_{i=1}^K p_i$ . To introduce the constraint (Section 2.5.1), we normalize the target function  $J$  by  $\sum_{i=1}^K p_i$ :

$$J' = \frac{\sum_{i=1}^K \log\left(\frac{p_i}{\sigma^2} \lambda_i + 1\right)}{\sum_{i=1}^K p_i}$$

Differentiating the above equation with  $p_i$ , the result is set to be zero as

$$\frac{\partial J'}{\partial p_i} = \frac{1}{\left(\sum_{i=1}^K p_i\right)} \frac{\frac{\lambda_i}{\sigma^2}}{\left(\frac{p_i \lambda_i}{\sigma^2} + 1\right)} - \frac{\sum_{i=1}^K \log\left(\frac{p_i}{\sigma^2} \lambda_i + 1\right)}{\left(\sum_{i=1}^K p_i\right)^2} = 0$$

For  $p_i$  fulfilling the above equation, we let

$$\sum_{i=1}^K \log\left(\frac{p_i}{\sigma^2} \lambda_i + 1\right) = C_m$$

Letting  $\sum_{i=1}^K p_i = P_x$ , we have

$$\frac{\lambda_i / \sigma^2}{(p_i \lambda_i / \sigma^2) + 1} P_x - C_m = 0$$

Thus, we get

$$p_i = \frac{P_x}{C_m} - \frac{\sigma^2}{\lambda_i} > 0 \quad (7.66)$$

The transmission of a data sequence using an eigenvector  $\mathbf{u}_i$  with  $p_i \leq 0$  is impossible. The above result means that a higher power is allocated for a data sequence with an eigenvector with a higher eigenvalue or a higher  $S/N$ . To get transmit power  $p_i$ , the constant  $P_x/C_m$  should be given. We can use an iterative method as follows.

The eigenvalues  $\lambda_i$  are assumed in a descending order as  $\lambda_1 \geq \lambda_2 \geq \lambda_3, \dots$

We take the summation of Equation 7.66 for  $p_i$  from  $i=1$  to  $r$  ( $r$  denotes the rank of matrix  $\mathbf{H}$ ):

$$P_x = \sum_{i=1}^r p_i = r\beta - \sum_{i=1}^r \frac{\sigma^2}{\lambda_i} \quad \left( \beta = \frac{P_x}{C_m} \right)$$

$$\text{Thus, } \beta = \frac{P_x}{C_m} = \left( P_x + \sum_{i=1}^r \frac{\sigma^2}{\lambda_i} \right) / r \text{ is given.}$$

Putting this value  $\beta$  into Equation 7.66, if we have  $p_i > 0$  ( $i=1, 2, \dots, r$ ), then  $p_i$  are decided with this  $\beta$ ; otherwise, similar procedure is taken by decreasing the number of the summation as  $r \rightarrow r-1$ . Repeating this procedure up to  $r-1$  times, the solution is always given. If the number  $K(<r)$  of multiplexed data sequences is given beforehand, the above procedure can be started from  $r=K$ .

The maximum value of the transmission capacity for  $K$  space-division multiplexed data sequences is given by applying  $\beta = \left( \frac{P_x}{C_m} \right) = \left( P_x + \sum_{i=1}^K \frac{\sigma^2}{\lambda_i} \right) / K$  and Equation 7.66 to Equation 7.65 as

$$J = \sum_{i=1}^K \log \left( \frac{\lambda_i}{\sigma^2} \left( \beta - \frac{\sigma^2}{\lambda_i} \right) + 1 \right) = \sum_{i=1}^K \log \left( \frac{\beta \lambda_i}{\sigma^2} \right)$$

Next, we take the minimum average bit error rate as the other criterion for power allocation. If transmit powers of each data sequence (substream) are denoted as  $p_k$  ( $k=1, 2, \dots, K$ ), average signal to noise power ratio for the sequences becomes  $\gamma_k = p_k \lambda_k / \sigma^2$ . Each sequence is transmitted using a modulation method with  $m_k$  bits per symbol. We denote by  $P_{ek}(\gamma_k)$  the bit error rates for a given modulation method. Then the average bit error rates for the given modulation method become  $m_k P_{ek}(\gamma_k)$ , since  $m_k$  bits are conveyed with a symbol. Then the average bit error rates  $\bar{P}_b$  for all data sequences become

$$\bar{P}_b = \frac{1}{N} \sum_{k=1}^K m_k P_{ek}(\gamma_k) = \frac{1}{N} \sum_{k=1}^K m_k P_{ek} \left( \frac{p_k \lambda_k}{\sigma^2} \right)$$

where  $N = \sum_{k=1}^K m_k$  denotes the total number of transmit bits for all the data sequences at a given symbol time. Our task is to determine power allocation  $p_k$  ( $k=1, 2, \dots, K$ )



**TABLE 7.1 Bit error rates for Gray coded modulations**

$P_e = \alpha \times \text{erfc}\left(\sqrt{\gamma_s/\beta}\right)$		
Modulation	$\alpha$	$\beta$
BPSK	1/2	1
QPSK	1/2	2
16-QAM	3/8	10
64QAM	7/24	42
256QAM	15/64	170

$\gamma_s = E_s/N_0$ ; symbol energy to noise power density ratio.

Approximate equation for QAM [17].

that makes  $\bar{P}_b$  to be minimum for a given total transmit power  $P_o = \sum_{k=1}^K p_k$ . The bit error rates with a given modulation/demodulation method are given by a complicated expression (Section 5.5.1). Here, we consider BPSK, QPSK, and M-QAM with the Gray coding and coherent detection. The error rates for these systems are given in Table 7.1. Even for these systems, it is difficult for us to solve the optimization problem. Therefore, we use the Chernoff bound [18] approximation as

$$Q(x) = \frac{1}{2} \text{erfc}\left(\frac{x}{\sqrt{2}}\right) \leq e^{-x^2/2}$$

Then we have  $P_{ek}(\gamma_k) \leq 2\alpha_k e^{-\gamma_k/\beta_k}$  and

$$\bar{P}_b \leq \frac{1}{N} \sum_{k=1}^K 2m_k \alpha_k e^{-\gamma_k/\beta_k} = \frac{1}{N} \sum_{k=1}^K 2m_k \alpha_k e^{-p'_k \lambda'_k / \beta_k}$$

where

$$p'_k = p_k/P_o, \quad \lambda'_k = \lambda_k \gamma_0 \quad (\gamma_0 = P_o/\sigma^2)$$

We use the Lagrangian multipliers method. The target function is expressed as

$$J = \frac{1}{N} \sum_{k=1}^K 2m_k \alpha_k e^{-p'_k \lambda'_k / \beta_k} + \mu \left( \sum_{k=1}^K p'_k - 1 \right)$$

where  $\mu$  is an unknown constant. Our task is to find  $p'_k (\geq 0)$  that makes the target function minimum. Letting  $\partial J / \partial p'_k = 0$ , we have

$$\frac{1}{N} 2m_k \alpha_k \left( \frac{-\lambda'_k}{\beta_k} \right) e^{-p'_k \lambda'_k / \beta_k} + \mu = 0$$

From this, we get

$$-\frac{p'_k \lambda'_k}{\beta_k} = \ln \frac{\mu}{2m_k \alpha_k \lambda'_k / (N \beta_k)}$$

Then  $p'_k (\geq 0)$  are given as

$$p'_k = \frac{\beta_k}{\lambda'_k} \left[ \ln \left( \frac{m_k \alpha_k \lambda'_k}{\beta_k} \right) - \xi \right]$$

where  $\xi = \ln(\mu N/2)$  and it is obtained with the condition  $\sum_{k=1}^K p'_k = 1 (p'_k \geq 0)$  through the iterative method mentioned earlier. The  $m_k$  bits per symbol are assumed beforehand in the above arguments. Actually, however, the combination of  $m_k$  for a given total number of  $N$  bits should be determined to get the minimum average error rate  $\bar{P}_b$ .

**Adaptive Signal Separation Methods.** In the methods for MIMO SDM discussed so far, the channel characteristics are known to separate interfering signals (sub-streams). Furthermore, the signal separation is carried out by a linear weighted sum of the signals at each antenna element (space filtering). Here, MIMO multiplexing methods without one of the above conditions are discussed. The linear weighting methods without knowledge of channel characteristics are described first. Then non-linear signal processing methods are discussed. Those methods are based on signal processing adaptive to channel characteristics.

Weighting coefficients in a MIMO antenna system can be determined without knowing the channel characteristics, if a reference (desired) signal is given; the weights are given by minimizing a target (cost) function, which is a function of an error defined as difference between the reference signal and the linear weighted sum of signals at each antenna element (Section 2.5). In a MIMO system, the error is caused by interference between multiplexed signals and noise. Since data and noise are random, the error becomes a stochastic variable and the target function is defined as an average value over the random variable.

A choice of the target function is the average of the absolute value of the error. The other one is the square of the former. Methods with the former and the latter ones are called zero-forcing (ZF) method and minimum mean square error (MMSE) method, respectively. The latter shows a superior performance over the former owing to its compromising effect between the interference and noise. The principles of the methods are described in Section 2.5.

Methods with the inverse matrix and diagonalization of the channel matrix with eigenvectors belong to the ZF method, since the result never reflects the effect of noise. The weighting coefficients appear at the transmit or receive side for the inverse matrix method and at both sides for the eigenvector method. These coefficients can be determined with the ZF method.

The method with diagonalization of correlation matrix ( $\mathbf{H}^H \mathbf{H}$ ) falls into the MMSE method, and it shows the best performance, since it achieves the transmit and receive MRC effect while nulling interference between the multiplexed signals. At the receiver, any method like maximal ratio combining, ZF, MMSE, and nonlinear method described later gives the same result as long as the signal processing is done at the transmitter side corresponding to the diagonalization of the correlation matrix.

The optimization of receive weights is carried out iteratively to minimize the target function without knowledge of the channel characteristics. However, the optimization of the transmit weights by an iterative method is impractical since signals should be transmitted in every step of the iteration. When channel characteristics were known, the transmit weights optimization with the iterative method becomes practical by assuming a virtual receiver at the transmit side.

The optimum method to suppress interference between substreams is to subtract interfering signals from a desired signal. If we apply this method to every substream, the substream signals are separated (multiuser receiver). The principle of interference suppression with a nonlinear signal processing is based on the subtraction of interfering signals replicated at the receiver side. The replication of signals at the receiver is possible for a digital signal, since it has a finite state and a finite length of sequence: we cannot apply the method to analog signals. The total number of combination of received signal at a given symbol time is  $L^K$  where  $K$  is the number of transmit data streams and  $L$  is the modulation levels, which is assumed the same for all the streams.

When noises at each receive antenna are independent and data occur equally likely, the most likely receive data are given by the maximum likelihood detection (MLD). The method is in essence the same as the maximum likelihood sequence estimation (MLSE; Section 3.3.8), except that the signal vectors are in space domain instead of time domain in MLSE. The algorithm is to find a candidate signal vector  $\hat{\mathbf{x}}_i$  minimizing the target function  $\|\mathbf{e}\|^2 = \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_i\|^2$  where  $\mathbf{e} = \mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_i$ ,  $\mathbf{y}$  is the received signal vector, and  $\mathbf{H}$  denotes the propagation matrix. In the following, the number of receive antenna elements is assumed the same as the number  $K$  of the data sequences for simplicity.

The number of candidate symbols for a symbol instance becomes  $L^K$ . Reduction of calculation complexity is a key issue for practical implementation of the algorithm. Some methods are known for the complexity reduction. All of the methods are based on a systematic search for candidate vectors without searching for all of the candidates (round-robin search). A representative one is QR decomposition of  $\mathbf{H}$  matrix as follows. A matrix  $\mathbf{H}$  is decomposed with a unitary matrix  $\mathbf{Q}$  and an upper triangular matrix  $\mathbf{R}$  as

$$\mathbf{H} = \mathbf{Q}\mathbf{R} \quad (7.67)$$

Then an error vector  $\mathbf{e}$  becomes

$$\mathbf{e} = \mathbf{y} - \mathbf{Q}\mathbf{R}\hat{\mathbf{x}}_i$$

Multiplying  $\mathbf{Q}^H$  to the above equation from the left side and using  $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}$  (unit matrix), we have

$$\mathbf{e}' = \mathbf{Q}^H\mathbf{e} = \mathbf{Q}^H\mathbf{y} - \mathbf{R}\hat{\mathbf{x}}_i \quad (7.68)$$

When a unitary matrix is applied to a vector, it never changes the norm of the vector, that is,  $\|\mathbf{e}\|^2 = \|\mathbf{Q}^H\mathbf{e}\|^2$ . Therefore, we may find  $\hat{\mathbf{x}}_i$ , which makes the norm of

the vector  $\mathbf{e}'$  to be minimum. Only the  $K$ th component  $\hat{x}_K$  of  $\hat{\mathbf{x}}_i$  contributes to  $|e'_K|^2$ , since  $R$  is a upper triangular matrix and the last ( $K$ th) component  $e'_K$  of  $\mathbf{e}'$  becomes  $e'_K = y'_K - r_{KK}\hat{x}_K$ , where  $y'_K$  is the  $K$ th component of the vector  $\mathbf{y}' = \mathbf{Q}^H \mathbf{y}$  and  $r_{KK}$  is the  $(K, K)$  component of the matrix  $R$ .

Even though we find a  $\hat{x}_K$  that makes  $|e'_K|^2$  minimum, this is not assured to be the optimum one, since  $\|\mathbf{e}'\|^2 = \sum_{k=1}^K |e'_k|^2$ . When a candidate  $\hat{x}_K$  is given, we have an expression as  $\|\mathbf{e}'\|^2 = |e'_K|^2 + \sum_{k=1}^{K-1} |e'_k|^2$ . Next, we consider  $e'_{K-1}$ . When  $\hat{x}_K$  and  $\hat{x}_{K-1}$  are given,  $|e'_{K-1}|^2$  can be calculated using  $e'_{K-1} = y_{K-1} - r_{KK-1}\hat{x}_{K-1} - r_{K-1K}\hat{x}_K$ . Then we have  $\|\mathbf{e}'\|^2 = |e'_K|^2 + |e'_{K-1}|^2 + \sum_{k=1}^{K-2} |e'_k|^2$ . Proceeding this treatment successively,  $\|\mathbf{e}'\|^2$  can be calculated hierarchically.

The reduction of computational complexity is done in every step of search by testing limited candidates selected on a given criterion and omitting calculation for the unselected candidates. The method is known as the sphere decoding [19]. This method selects candidates of  $\hat{x}_k$  that meet  $\|\mathbf{e}'\|^2 \leq C_0$  ( $C_0$  is a given constant) in every step and finds a candidate  $\hat{\mathbf{x}} = (\hat{x}_K, \hat{x}_{K-1}, \dots, \hat{x}_1)$  that gives the minimum  $\|\mathbf{e}'\|^2$ . In the first step, more than one candidate meets  $|e'_K|^2 \leq C_0$  from  $\hat{x}_K$ , which takes  $L$  different states. In the second step, more than one  $\hat{x}_{K-1}$  is selected, which meets  $|e'_K|^2 + |e'_{K-1}|^2 \leq C_0$  for each of  $\hat{x}_K$  selected in the first step. The similar procedure to select candidates in the step is carried out until reaching the final step  $K$ . A candidate  $\hat{\mathbf{x}} = (\hat{x}_K, \hat{x}_{K-1}, \dots, \hat{x}_1)$  is fixed, which gives the minimum  $\|\mathbf{e}'\|^2$ .

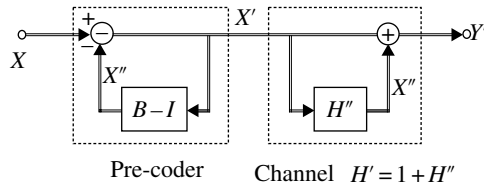
Another method is called the  $M$  algorithm [20]. This method selects  $M (< L)$  candidates from  $L$  states in every step on selection criteria where the candidates give smaller  $\delta_i = \sum_{k=i}^K |e'_k|^2$  ( $i=K, K-1, \dots, 1$ ) in step  $i$ .

The other method decomposes  $\mathbf{H}$  matrix into small matrices to select candidates with use of space filtering technique [21].

**Multiuser MIMO.** In the arguments so far, signal processing to separate signals is performed using all of the output signals from each receive antenna. Here, we consider a case when in downlink signals to multiple users (terminals) are space division multiplexed. Then a user terminal cannot use received signals at other terminals. In this situation, only a method is possible, which suppresses interference between the signal at the transmit side. For example, we consider two user signals are multiplexed with two antennas at base station and one antenna at a user terminal. Then received signals  $y_i$  for user  $i(i=1, 2)$  are given with transmit signals  $x_i$  and  $n_i$  noises as

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}$$

where  $h_{ij}$  are elements of propagation matrix. We use a notation,  $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ . Multiplying the inverse matrix of  $\mathbf{H}$  at the transmitter, we have the received signals  $\mathbf{y}'$  as  $\mathbf{y}' = \mathbf{H}\mathbf{H}^{-1}\mathbf{x} + \mathbf{n} = \mathbf{x} + \mathbf{n}$ .



**FIGURE 7.16** A system with inverse circuit of channel placed at the transmitter.

The received signal interference is removed. Since constant value multiplication to the signal and noise at a receive antenna never affects performance, we modify the propagation matrix as

$$\mathbf{H}' = \begin{pmatrix} 1 & h'_{12} \\ h'_{21} & 1 \end{pmatrix} \quad h'_{12} = \frac{h_{12}}{h_{11}}, \quad h'_{21} = \frac{h_{21}}{h_{22}}$$

If we set

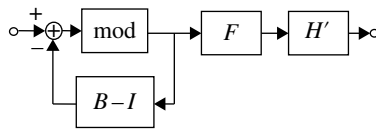
$$\mathbf{H}'' = \mathbf{H}' - \mathbf{I}$$

then the transmitter with the inverse matrix operation (precoding) is shown in Figure 7.16. Putting  $\mathbf{B} - \mathbf{I} = \mathbf{H}'' = \mathbf{H}' - \mathbf{I}$ , that is,  $\mathbf{B} = \mathbf{H}' = \mathbf{I} + \mathbf{H}''$ , the fact that  $\mathbf{y}' = \mathbf{x}$  is easily seen since  $\mathbf{y}'$  is produced from  $\mathbf{x}$  subtracted first and added next with  $\mathbf{x}''$ .

Difficulties of the precoding are the inherent instability due to the feedback circuit and increase of transmit power resulting into performance degradation compared with the ideal precoding known as the dirty paper coding (DPC), which gives the same channel capacity as a system without the intersignal interference [22]. Since DPC performance is achievable only infinitesimally, practical approaches have been investigated. A representative one is called the Tomlinson–Harashima precoding [23, 24]. The method is proposed as a preequalizer at the transmit side to suppress intersymbol interference in a channel as described in Section 7.3.6.

This method is applicable to a multiuser MIMO system, since the mathematical expression of intersignal interference becomes the same for both systems except for the difference between the expressions in time and space domains. However, only a precoding method is applicable for a multiuser MIMO system since received signals are not shared between different user terminals, while signal equalization in time domain can be carried out at both transmit and receive sides. The precoding method adopts a nonlinear signal processing, that is, modulo operation, which contributes to mitigate the inherent instability of a feedback system and the increase of transmit power. When the modulo operation is introduced to a MIMO system, intersignal interference should be suppressed in a successive manner [25] (time-domain equalization can be carried out starting from the first signal owing to the causality of a channel). To do this, we decompose the propagation matrix  $\mathbf{H}'$  as

$$\mathbf{H}' = \mathbf{S}\mathbf{F}^H$$



**FIGURE 7.17** Multuser MIMO system with Tomlinson–Harashima precoding.

where  $\mathbf{F}$  is a unitary matrix and  $\mathbf{S}$  is a lower triangular matrix. Figure 7.17 shows a multuser MIMO system with the Tomlinson–Harashima precoding. If we assume a new propagation matrix made of a cascade connection of matrices  $\mathbf{F}$  and  $\mathbf{H}'$ , the inverse matrix is given by letting  $\mathbf{B} = \mathbf{H}'\mathbf{F}$  when the modulo circuits are omitted. Then, we have

$$\mathbf{B} = \mathbf{S}\mathbf{F}^H\mathbf{F} = \mathbf{S} \quad (\mathbf{F}^H\mathbf{F} = \mathbf{I})$$

Since matrix  $\mathbf{S}$  is a lower triangular matrix, for a  $2 \times 2$  matrix as the above example, we get the following input–output relation expression for a channel with  $\mathbf{H}'\mathbf{F} = \mathbf{S}$ :

$$\begin{pmatrix} y'_1 \\ y'_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ s_{21} & 1 \end{pmatrix} \begin{pmatrix} x''_1 \\ x''_2 \end{pmatrix}$$

From this, we see no interference from signal  $x''_2$  to signal  $x''_1$ .

Similarly, for an  $N \times N$  matrix, signal  $x''_i$  has no interference from signals  $x''_j (j > i)$ . Thus the interferences in the circuit  $\mathbf{S}$  are cancelled by subtracting the interferences at the precoder as

$$x''_i = x_i - \sum_{j=1}^{i-1} s_{ij}x''_j$$

For a multuser MIMO system with modulo circuits, signals  $x''_2, \dots, x''_N$  are given successively starting from signal  $x''_1$ . The modulo operation is expressed as

$$a' = a \bmod(2A) = a + 2mA \quad (m = 0, \pm 1, \pm 2, \dots)$$

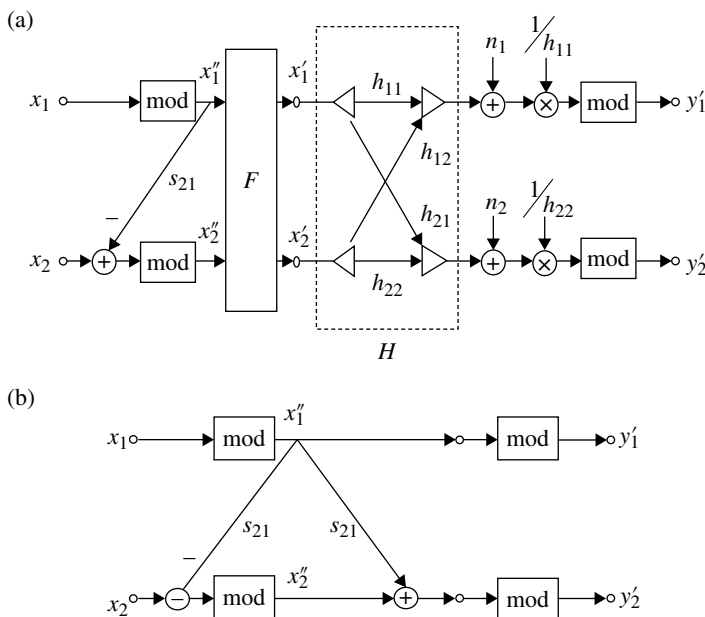
where integers  $m$  are determined so that the levels of signal  $a'$  fall in a range of  $-A \sim A$ . Signal levels of signal  $x_i$  are assumed in the same range. Figure 7.18 shows an example system for a two-user MIMO system with two transmit antenna at a base station. As seen from Figure 7.18b, we have

$$\begin{aligned} x''_2 &= (x_2 - s_{21}x''_1) \bmod(2A) \\ &= x_2 - s_{21}x''_1 + 2mA \end{aligned}$$

Therefore, we get

$$\begin{aligned} y'_2 &= (x''_2 + s_{21}x''_1 + 2mA) \bmod(2A) \\ &= (x_2 + 2mA) \bmod(2A) \\ &= x_2 \end{aligned}$$

Since  $\mathbf{F}$  is a unitary matrix, the transmit power never changes with it.



**FIGURE 7.18** Two-user MIMO system with Tomlinson–Harashima precoding. (a) Actual system and (b) equivalent circuit ( $n_1=n_2=0$ ).

### 7.3 ADAPTIVE AUTOMATIC EQUALIZER

As the signal bandwidth becomes comparable to the coherence bandwidth of the fading mobile radio channel (frequency-selective fading), the signal suffers from waveform distortion (Section 4.4). The automatic equalizer is known to remove distortion at the receiver. Automatic equalizer technologies are well developed and are applied to data transmission through public telephone channels. Many of the techniques can be applied to mobile radio channels. However, the mobile radio has a special requirement for the equalizer: adapting to the fast fading channel. Research and development on equalizers is continuing at the time of writing. Here, a short introductory description of automatic equalizers and applications to mobile radio channels is given.

The most common example of signal distortion in radio channels is the so-called ghost on a (analog) TV picture. This phenomenon occurs when echo signals with significant power and delay are received after the main signal. There are ways to resolve the ghost problem: (i) improve the receiving antenna system, for example, the directivity; (ii) enter a CATV system; and (iii) use an echo canceller. The echo canceller is a type of equalizer, which we will discuss later. Digital signal transmission has an advantage over analog signal transmission in view of equalizing the channel because the digital signal takes discrete values, as will be shown in the following.

### 7.3.1 Linear Equalizer

A distorted signal can be equalized by using a filter that has a transfer function that is the inverse of the channel transfer function [26–28]. For example, consider a two-path channel with impulse response  $h(t)$  given as

$$h(t) = A_0\delta(t) + A_1\delta(t - t_1) \quad (7.69)$$

The transfer function becomes

$$H(\omega) = A_0 + A_1e^{-j\omega t_1} \quad (7.70)$$

and for the inverse filter  $H^{-1}(\omega) = 1/H(\omega)$ .

Since digital signal processing is generally adopted for automatic equalizers, it is convenient to use a discrete-time (sampled) representation of the received signal. The received signal sampled with sampling period of  $T$  is given as

$$\begin{aligned} r(nT) &= s(t) * h(t) \big|_{t=nT} \\ &= A_0s(nT) + A_1s(nT - t_1) \quad (n = 0, \pm 1, \pm 2, \dots) \end{aligned}$$

where  $s(t)$  denotes transmit signal and the second term of the right-hand side represents the echo signal. For simplicity, we assume  $t_1 = T$  in the following. Denote a  $T$ -time delay element by  $z^{-1}$  ( $z$ -transform). The channel transfer function is then expressed as

$$H(z) = A_0 + A_1z^{-1}$$

where  $z = e^{j\omega T}$ .

The transfer function of the inverse filter is

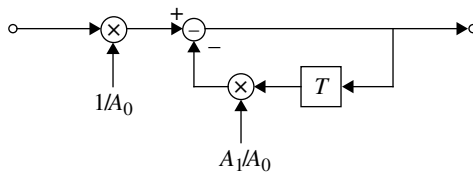
$$\begin{aligned} H(z)^{-1} &= \frac{1}{A_0 + A_1z^{-1}} \\ &= \frac{1}{A_0} \frac{1}{1 + \left(\frac{A_1}{A_0}\right)z^{-1}} \end{aligned} \quad (7.71)$$

The above transfer function is realized with the feedback circuit in Figure 7.19. We can expand Equation 7.71 as

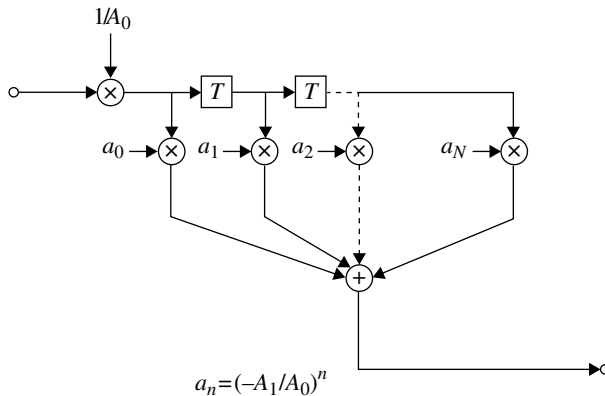
$$H(z)^{-1} = \frac{1}{A_0} \sum_{m=0}^{\infty} a^m z^{-m}$$

where  $a = -A_1/A_0$ . This transfer function can be realized by the circuit known as the transversal filter as in Figure 7.20. The impulse response converges if  $|az| = |a| < 1$ ;





**FIGURE 7.19** Negative feedback equalizer for a two-path channel.



**FIGURE 7.20** Transversal equalizer.

otherwise, it diverges. The results correspond to whether the feedback circuit is stable or not. We rewrite Equation 7.71 as

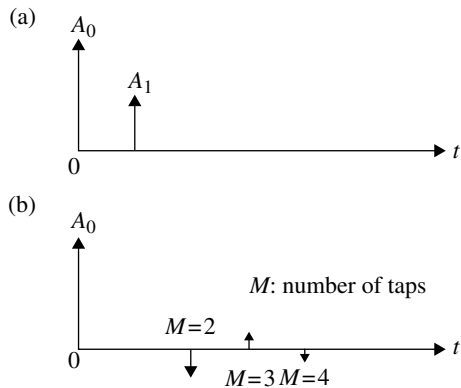
$$H(z)^{-1} = \frac{1}{A_1} \frac{z}{1 + (A_0 / A_1)z} = \frac{1}{A_1} z \sum_{m=0}^{\infty} b^m z^m$$

where  $b = -A_0 / A_1$ . The circuit is stable when  $|b| < 1$ . Replacing  $z^{-1}$  by  $z$  in a filter means the time reversal.

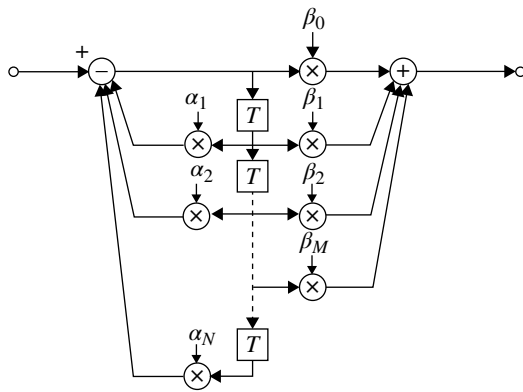
When the transversal equalizer is implemented, we must truncate the length of the filter. To show the effect of the truncation, an example of the impulse response of the system (channel plus equalizer) is shown in Figure 7.21. For the system, the channel impulse response as a whole becomes  $H^{-1}(z) = 1 - (az^{-1})^M$  ( $M$ : number of taps).

We can generalize the equalizer with finite length as shown in Figure 7.22. This structure is called a linear equalizer since it is realized by a linear circuit. The transfer function becomes

$$H(z) = \frac{N(z^{-1})}{D(z^{-1})} = \frac{\sum_{m=0}^M b_m z^{-m}}{1 + \sum_{n=0}^N a_n z^{-n}}$$



**FIGURE 7.21** Effect of transversal equalizer for impulse response ( $A_1/A_0=0.5$ ). (a) Channel and (b) channel + equalizer.



**FIGURE 7.22** Linear equalizer.

For the stability of the circuit, the equation  $D(z^{-1})=0$  should have roots, all of whose absolute values are less than 1 in the  $z$ -plane (Section 2.4.4). The task of designing a linear equalizer becomes one of assigning coefficients  $a_n$  and  $b_m$  to get the best performance under the constraints of stable operation and size of the equalizer. Because of the instability problem, linear equalizers exclusively use the transversal filter structure. The criterion for equalizer performance should be error rate. However, this is a difficult problem since the error rate is a highly nonlinear function of the tap coefficients. Therefore, the error, which is the difference between the desired signal and the equalized one, is considered as a performance criterion.

For digital transmission systems, we are interested in the signal errors at the sampling or decision-making instants. In this case, the equalizer can be operated on the

basis of one sample per symbol (symbol rate sampling equalization). However, it is shown that symbol rate sampling equalization is sensitive to the sample timing error: the performance of the equalizer deteriorates with the timing error. Hence, joint optimization for the sample timing and equalization must be carried out. Double sampling per symbol resolves this problem [27].

### 7.3.2 Performance Criteria for Equalization

Let us consider a digital transmission system that has no intersymbol interference when channel distortion is absent. The sampled signal becomes  $a_k (k = 0, 1, 2, \dots)$ , where  $a_k$  takes a value from a discrete set of levels corresponding to the transmitted data. Intersymbol interference deviates the sample signal from  $a_k$  due to channel distortion. We denote  $y(t)$  as an input signal to an equalizer:

$$y(t) = \sum_k a_k h(t - kT)$$

where  $h(t)$  denotes the impulse response for the tandem combination of a transmit filter, the channel, and a receive filter. The error or the intersymbol interference at  $t = kT$  is given by  $y(kT) - a_k$ .

Let us consider a transversal equalizer. The output signal of the equalizer becomes

$$z(kT) = \sum_{n=-N}^N y(kT - nT') w_n \quad (T' \leq T)$$

where  $w_n$  are the tap coefficients and  $T'$  is the sampling period for the equalizer. Hereafter, we use the notation  $z_k = z(kT)$  and  $y_k = y(kT)$ . The error is given by

$$e_k = z_k - a_k$$

We have two criteria to measure the error: mean squared error  $\langle e_k^2 \rangle$  and the maximum absolute error  $\max |e_k|$ . We must adjust  $w_n$  to minimize  $\langle e_k^2 \rangle$  or  $\max |e_k|$ .

#### Mean Square Error Criterion and Mean Square Algorithm

Assuming symbol rate equalization ( $T' = T$ ), we have

$$\langle e_k^2 \rangle = \left\langle \left( \sum_{n=-N}^N y_{k-n} w_n - a_k \right)^2 \right\rangle$$

In order to minimize  $\langle e_k^2 \rangle$ , we require the derivatives of  $\langle e_k^2 \rangle$  with respect to  $w_n$  to be zero. We then have  $2N + 1$  linear equations given as

$$\frac{\partial}{\partial w_n} \langle e_k^2 \rangle = 2 y_{k-n} \left\langle \left( \sum_{m=-N}^N y_{k-m} w_m - a_k \right) \right\rangle = 0 \quad (n = -N, -N + 1, \dots, N)$$

The above equations can be rewritten in matrix form as

$$[Y_{mn}] \begin{bmatrix} w_{-N} \\ w_{-N+1} \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} c_{-N} \\ c_{-N+1} \\ \vdots \\ c_N \end{bmatrix} \quad (7.72)$$

where the matrix elements are given as

$$Y_{mn} = \langle y_{k-m} y_{k-n} \rangle$$

and

$$C_m = \langle y_{k-m} a_k \rangle$$

Multiplying Equation 7.72 by the inverse matrix  $[Y_{mn}]^{-1}$ , we have the optimum tap coefficients given as

$$\begin{bmatrix} w_{-N} \\ w_{-N+1} \\ \vdots \\ w_N \end{bmatrix} = [Y_{mn}]^{-1} \begin{bmatrix} c_{-N} \\ c_{-N+1} \\ \vdots \\ c_N \end{bmatrix}$$

In the above argument, we must know  $a_k$ , which are the transmitted data. For this purpose, a given sequence of  $a_k$ , called a test signal, reference signal, or training signal, is transmitted prior to the information signal. In practice, real-time matrix inversion is difficult. Alternatively, an iterative method may be applied. Notice that  $(\partial/\partial w_n) \langle e_k^2 \rangle$  tells us whether the mean square error  $\langle e_k^2 \rangle$  increases  $((\partial/\partial w_n) \langle e_k^2 \rangle > 0)$  or decreases  $((\partial/\partial w_n) \langle e_k^2 \rangle < 0)$  when the tap coefficient  $w_n$  is changed to value away from the optimum value. Thus, we can know whether we should increase or decrease  $w_n$  to get closer to the optimum value. The tap coefficients are updated as

$$w_n(j+1) = w_n(j) - \Delta \frac{\partial}{\partial w_n} \langle e_k^2(j) \rangle \quad (j=0,1,2, \dots)$$

where  $j$  denotes the  $j$ th iteration and  $\Delta$  is a small constant. This method is known as the steepest descent method. Using the expression

$$\begin{aligned} \frac{\partial}{\partial w_n} \langle e_k^2 \rangle &= \left\langle y_{k-n} \left( \sum_{m=-N}^N y_{k-m} w_m - a_k \right) \right\rangle \\ &= \langle y_{k-n} e_k \rangle \end{aligned}$$

we have the least mean square (LMS) automatic equalizer as in Figure 7.23 where the mean value  $= \langle y_{k-n} e_k \rangle$  is replaced with the instantaneous value.

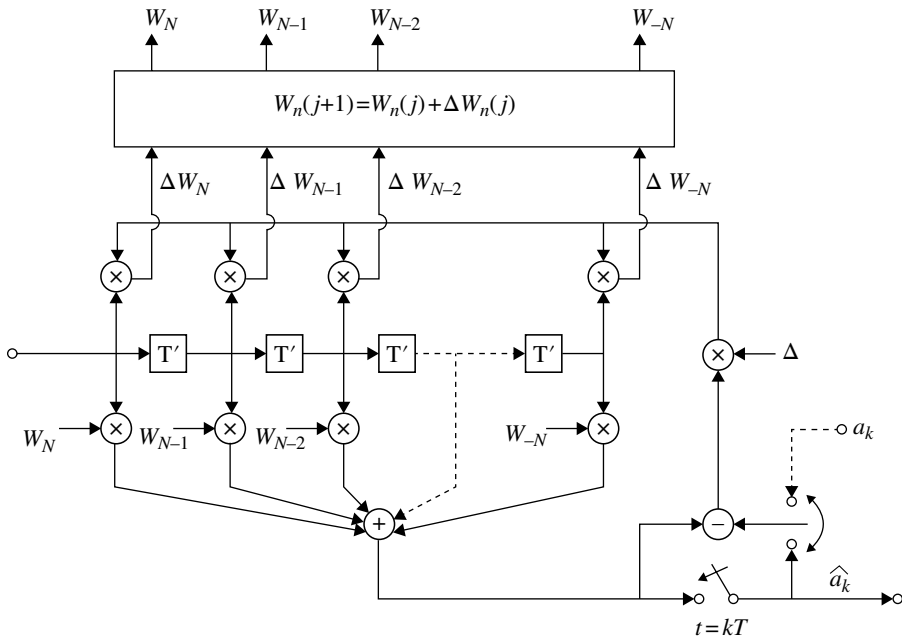


FIGURE 7.23 Least mean square equalizer.

In the training mode of operation, the reference signal  $a_k$  is used, while in the information signal transmit mode of operation, the decided data  $\hat{a}_k$  are used. If the channel characteristics never change, the tap coefficients  $w_n$  are fixed after the training is finished. For time-varying channels such as mobile radio channels, the equalizer adapts to the time-varying channel with an adaptive algorithm using the decided data.

**Peak Distortion Criteria and ZF Algorithm.** Peak distortion is defined as

$$\begin{aligned}
 C_1 &= \max_{\{a_k\}} |z_k - a_k| \\
 &= \max_{\{a_k\}} |y_{k-n} w_n - a_k| \\
 &= \max_{\{a_k\}} \left| \sum_n a_{k-n} t_n - a_k \right|,
 \end{aligned}$$

where

$$t_n = \sum_{i=-N}^N h_{n-i} w_i$$

and  $h_n = h(nT)$ .  $t_n$  is the sampled impulse response of the total system including the equalizer. For simplicity, assume  $\max |a_k| = 1$  and  $t_0 = 1$ . We then have

$$C_1 = \sum_{n(\neq 0)} |t_n|$$

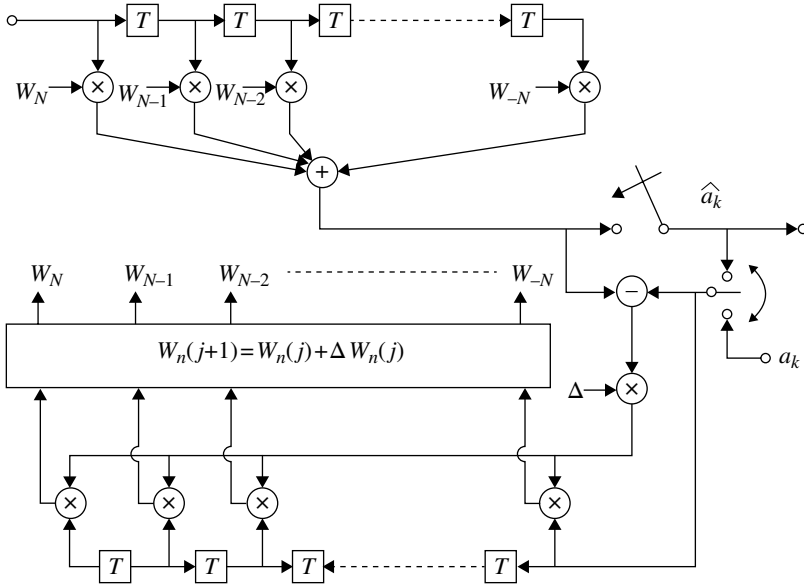


FIGURE 7.24 Zero-forcing equalizer.

$C_1$  is minimum ( $=0$ ) if  $t_n = 0$  ( $n \neq 0$ ) are forced to be zero.  $t_n = 0$  ( $n \neq 0$ ) means we have no intersymbol interference at the output of the equalizer. It is shown in [27] that if  $\sum_{i(\neq 0)} |h_i| < |h_0|$ ,  $t_n = 0$  ( $n \neq 0$ ) can be achieved by adjusting the tap coefficient  $w_n$  so that

$$\langle a_{k-n}(z_k - a_k) \rangle = 0 \quad (n = -N, \dots, N) \quad (7.73)$$

Since  $a_k$  are random values, we rewrite Equation 7.73 as

$$\langle a_{k-n} \left( \sum_m a_{k-m} t_m - a_k \right) \rangle = \langle a_{k-n}^2 \rangle t_n \quad (n \neq 0)$$

The recursive algorithm to achieve Equation 7.73 is

$$w_n(j+1) = w_n(j) - \Delta a_{k-n}(z_k - a_k) \quad (n = -N, \dots, N)$$

where  $\Delta$  is a small constant. An equalizer with the ZF algorithm is shown in Figure 7.24. The ZF algorithm yields  $\langle a_{k-n} e_k \rangle = 0$ , which is different from the LMS algorithm, which yields  $\langle y_{k-n} e_k \rangle = 0$ .

The linear equalizer with either the LMS or ZF algorithm is an inverse filter for the channel as long as noise is absent. (If we consider the noise, the difference between the two algorithms is seen: the LMS algorithm tends to find optimum tap coefficients, compromising between the effects of intersymbol interference and noise power increase, while the ZF algorithm never takes account of noise.) Thus, the

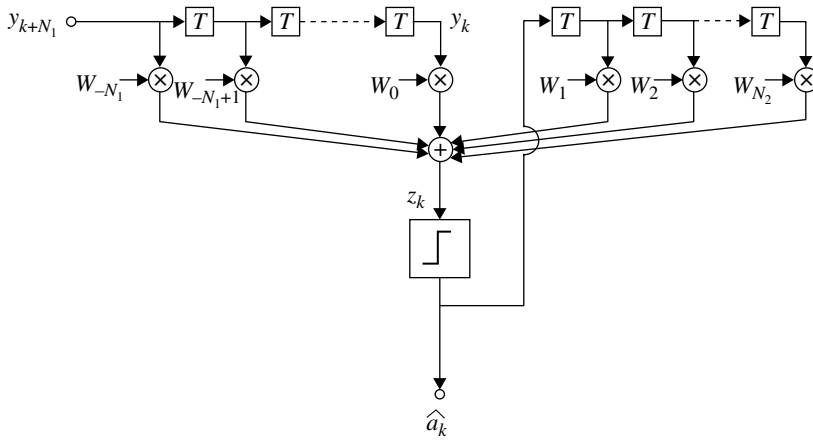


FIGURE 7.25 Decision feedback equalizer.

equalization of the channel is equivalent to prediction of the channel characteristics. From this viewpoint, the linear equalizer principle plays an important role in adaptive equalization. Linear equalizers are analyzed in [26, 27].

### 7.3.3 Decision Feedback Equalizer

Figure 7.25 depicts the decision feedback equalizer [26]. The equalizer output signal,  $z_k$ , can be expressed as

$$z_k = \sum_{n=-N_1}^0 y_{k-n} w_n + \sum_{n=1}^{N_2} \hat{a}_{k-n} w_n \quad (7.74)$$

The decision feedback equalizer differs from the linear feedback equalizer in that decided data  $\hat{a}_k$  are fed back instead of  $z_k$ . Since a nonlinear element, namely, decision circuit, is introduced in the feedback path, we have no instability problem.

For a performance criterion, we can use either the LMS error or the peak distortion. The mean square error criterion is used in most systems. The target function is then  $\langle |e_k|^2 \rangle = \langle |z_k - \hat{a}_k|^2 \rangle$ . The analysis of the system is difficult when taking decision errors into consideration. We therefore assume no decision errors:  $\hat{a}_k = a_k$ . Furthermore, we assume random data. Using Equation 7.74, making the derivative of  $\langle |e_k|^2 \rangle$  with respect to the feedback coefficients  $w_n$  zero, we can show that the feedback coefficients  $w_n$  are expressed in terms of the feed-forward coefficients as

$$w_n = - \sum_{k=-N_1}^0 w_k h_{n-k} \quad (n=1, 2, \dots, N_2)$$

If we assume  $h_i = 0$  for  $i < 0$  or  $i > L$ , then  $w_n = 0$  for  $n > L$ .

Making the derivative of  $\langle |e_k|^2 \rangle$  with respect to the feed-forward coefficients zero and using the above equation, we have

$$\sum_{k=-N_1}^0 \varphi_{nk} w_k = h_{-n}^* (n = -N_1, \dots, 1, 0)$$

where

$$\varphi_{nk} = \sum_{m=0}^{-n} h_m^* h_{n+m-k} (n, k = -N_1, \dots, 1, 0)$$

If we use the steepest descent method, the tap coefficients  $w_n$  are updated in the feed-forward sections as

$$w_n(j+1) = w_n(j) - \Delta e_j y_{j-n}^* (n = -N_1, \dots, -2, -1, 0)$$

and in the feedback sections as

$$w_n(j+1) = w_n(j) - \Delta e_j \hat{a}_{j-n} \quad (n = 1, 2, \dots, N_2)$$

where  $\Delta$  is a small constant.

### 7.3.4 The Viterbi Equalizer

The MLSE receiver gives the optimum performance for a digital communication system where intersymbol interference exists. The Viterbi algorithm (Section 3.3.8) is exclusively used in MLSE receiver. We can apply this method to frequency-selective fading channels. For this purpose, an estimate of the time-varying channel impulse response is required for the design of the matched filter. The Viterbi equalizer is shown in Figure 7.26.

The Viterbi equalizer compensates for the channel distortion in a way that is different in the usual sense of equalizers: it never produces a circuit with inverse transfer function of the channel, but replica signals at a receiver. Nevertheless, it is the optimum receiver as long as the channel impulse response is known. Like the DFE, nonlinear operations are made on the received signal, so it is also a nonlinear equalizer.

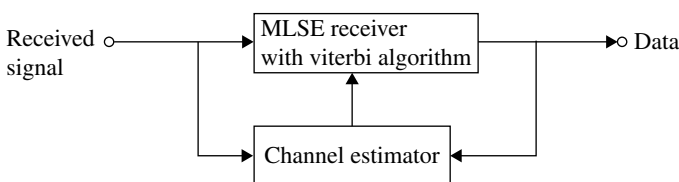


FIGURE 7.26 Viterbi equalizer.



### 7.3.5 Adaptation and Prediction Algorithm

Besides the ZF and LMS error algorithms, some other methods are known, such as recursive least squares (Section 2.5.2) or the Kalman algorithms, the fast Kalman, and lattice algorithms. Most of the equalizers assume a training sequence for updating the tap coefficients. A category of equalizers that requires no training sequence is known as blind equalizers [29]. Description of these topics is beyond the scope of this book. Readers can refer to [26–31] for more detail.

### 7.3.6 Preequalization

The arguments so far assumed equalization at a receive side. Here, we discuss the Tomlinson–Harashima [23, 24] precoding, where equalization is done at a transmit side, and the vector coding [32], where equalization is carried out at both the transmit and receive sides. It is understood that channel characteristics are known at the transmit side.

Consider a time-discrete system and denote a transmit signal, a received signal, and noise by  $x[n] = x_n$ ,  $y[n] = y_n$ , and  $z[n] = z_n$ , respectively. Then we have

$$y_n = \sum_{m=1}^N h_{n-m} x_m + z_n \quad (n = 1, 2, \dots, N + M)$$

where  $h_n = h[n]$  are the channel impulse response, which is assumed as

$$h_i = 0 \quad (i < 0, i > M)$$

Letting  $h_0 = 1$  never loses generality, since the signal to noise power ratio remains the same when  $y_n$  are multiplied by  $1/h_0$ .

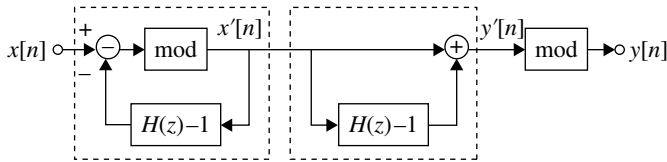
**The Tomlinson–Harashima Preequalization (Precoding).** A system with this method is shown in Figure 7.27. We need not assume a finite signal length, since an input signal can be applied continuously. In this figure,  $H(z)$  is the z-transform of the channel impulse response  $h[n](=h_n)$ , that is,

$$H(z) = 1 + h_1 z^{-1} + h_2 z^{-2} + \dots + h_M z^{-M}$$

To understand this system, the modulo operation is omitted first for simplicity. Then the transfer function  $P(z)$  of the preequalizer becomes

$$P(z) = \frac{1}{1 + (H(z) - 1)} = \frac{1}{H(z)}$$

Therefore, we have  $P(z) = H^{-1}(z)$  or  $P(z)H(z) = 1$ : the preequalizer is the inverse circuit of the channel and intersymbol interference is removed. This preequalizer is a feedback circuit, which inherently shows difficulty that the system tends to be unstable or increases transmit power. The Tomlinson–Harashima precoding avoids this problem by introducing a nonlinear signal processing, namely, the modulo operation.



**FIGURE 7.27** Transmission system with Tomlinson–Harashima precoding.

If we let  $h_1 = 1, h_i = 0 (i \geq 2)$ , the channel becomes equivalent to the duobinary encoding and the preequalizer is the same as the precoder for the duobinary signaling system (Section 3.2.6). The Tomlinson–Harashima precoding can cope with any intersymbol interference ( $h_i$ ) and can be applied to analog signal also.

A modulo circuit outputs a signal  $x'$  for an input signal  $x$  as

$$x' = x(\text{mod} 2A) = x + 2A \cdot i$$

where  $A$  is an arbitrary constant and an integer  $i$  is chosen so that the level of signal  $x'$  falls in the range of  $-A \sim A$ . The output signal level of the precoder (a feedback circuit) is limited within  $|x'| < A$  and the inherent instability is avoided.

If  $|x| < A$ , we have

$$\begin{aligned} y'[n] &= x[n](\text{mod} 2A) \\ &= x[n] + 2Ai \end{aligned}$$

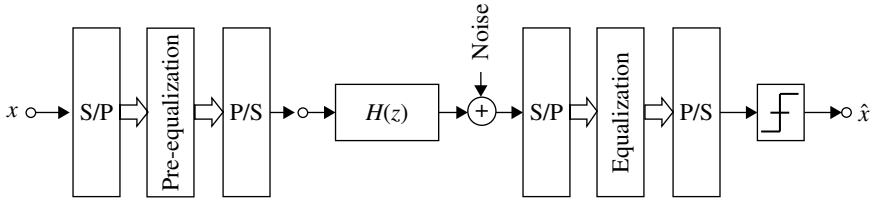
Performing modulo operation for the received signal gives

$$y[n] = y'[n](\text{mod} 2A) = x[n]$$

The Tomlinson–Harashima precoding has not been given much attention since channel characteristics are needed at the transmit side and performance advantage is not much high against (nonlinear) equalization at the receive side. Now, the method is indispensable for multiuser MIMO systems (Section 7.2.3) where signal processing is impossible at the receiver side.

In a digital transmission system, the parameter  $A$  of modulo operation should be given taking error rate performance into consideration; for an  $M$ -level QAM system with its inphase and quadrature-phase digital levels of  $\pm B, \pm 3B, \pm 5B, \dots, \pm(\sqrt{M}-1)B$ ,  $A = \sqrt{M}B$  is taken.

**Vector Coding.** Figure 7.28 shows a block diagram of the system. In this system, intersymbol interference due to channel dispersion is removed with linear signal processing at both transmit and receive sides, while a matched filter is used for each data symbol [32]. The mathematical treatment of the vector coding is equivalent that for an eigen-beam MIMO system (Section 7.2.3), except for the difference in signal representations, namely, in time domain and space domain.



**FIGURE 7.28** Transmission system with vector coding.

We use a column vector expression of a data block as  $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$  (T: transpose).

Preequalizing is performed with a linear signal processing as

$$x_n' = \sum_{m=1}^N p_{nm} x_m$$

where  $p_{nm}$  are coefficients, which will be given corresponding to channel characteristics later. Using matrix notations, we have

$$\mathbf{x}' = \mathbf{P}\mathbf{x} \quad (7.75)$$

where  $\mathbf{P}$  is a matrix with elements  $p_{nm}$ .

Pulse shaping filters with the square root Nyquist roll-off characteristics are assumed at transmit and receive sides. Neglecting the noise, received signals sampled with symbol time duration at the output of the receive filter are given as

$$y_n = \sum_{m=1}^N h_{n-m} x_m' \quad (n = 1, 2, \dots, N + M)$$

where  $M$  is the length of channel impulse response. A matrix expression of the above equation becomes

$$\mathbf{y} = \mathbf{H}\mathbf{x}' \quad (7.76)$$

where

$$H = \begin{pmatrix} h_0 & 0 & 0 & \cdots & 0 \\ h_1 & h_0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_M & h_{M-1} & h_{M-2} & \cdots & 0 \\ 0 & h_M & h_{M-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & h_M \end{pmatrix} \quad (N \times (N + M))$$

We use a linear equalizer at the receive side. The equalizer output  $y'_n (n = 1, 2, \dots, N)$  is given as

$$\begin{aligned} y'_n &= \sum_{m=1}^{N+M} q_{nm} (y_m + z_m) \\ &= \sum_{m=1}^{N+M} q_{nm} y_m + \sum_{m=1}^{N+M} q_{nm} z_m \quad (n = 1, 2, \dots, N) \end{aligned}$$

where  $q_{nm}$  are coefficients of the receive linear filter (equalizer) and  $z_n (n = 1, 2, \dots, N + M)$  are the sampled noise. Using a matrix notation, we have

$$\mathbf{y}' = \mathbf{Q}\mathbf{y} + \mathbf{Q}\mathbf{z} \quad \text{where} \quad \mathbf{Q} = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1N+M} \\ q_{21} & q_{22} & \cdots & q_{2N+M} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N1} & q_{N2} & \cdots & q_{NN+M} \end{pmatrix}$$

Inserting Equations 7.75 and 7.76 into the above equation, we have

$$\mathbf{y}' = \mathbf{QHP}\mathbf{x} + \mathbf{Q}\mathbf{z}$$

The first and second terms express signals and noises, respectively. We denote these as

$$\mathbf{s}' = \mathbf{QHP}\mathbf{x}$$

$$\mathbf{n}' = \mathbf{Q}\mathbf{z}$$

Our target for the present is to get  $\mathbf{P}$  and  $\mathbf{Q}$ , which give the maximum signal to noise power ratio under a constraint of a fixed transmit power. We use notation  $\|\cdot\|^2$ , which means the squared norm, for example,  $\|\mathbf{x}\|^2 = \sum_{n=1}^N |x_n|^2$ .

The average energy of a transmit signal  $P'_s$  becomes

$$\begin{aligned} P'_s &= \langle \|\mathbf{x}'\|^2 \rangle = \langle \|\mathbf{P}\mathbf{x}\|^2 \rangle = \langle (\mathbf{P}\mathbf{x})^H \cdot \mathbf{P}\mathbf{x} \rangle \\ &= \left\langle \sum_{n=1}^N \left| \sum_{m=1}^N p_{nm} x_m \right|^2 \right\rangle = \left\langle \sum_{n=1}^N \left( \sum_{m=1}^N p_{nm} x_m \right) \left( \sum_{m'=1}^N p_{nm'}^* x_{m'}^* \right) \right\rangle \end{aligned} \quad (7.77)$$

where we assume transmit signals are independent and have the following property:

$$\langle x_m^* x_n \rangle = \begin{cases} P_s & (m = n) \\ 0 & (m \neq n) \end{cases} \quad (m, n = 1, 2, \dots, N)$$

Then Equation 7.77 becomes

$$P'_s = \sum_{n=1}^N \sum_{m=1}^N |p_{nm}|^2 P_s \quad (7.78)$$

For noises, we assume

$$\langle z_m^* z_n \rangle = \begin{cases} \sigma^2 & (m = n) \\ 0 & (m \neq n) \end{cases}$$

To keep transmit signal power per symbol to be constant, the following normalization for  $p_n = (p_{n1}, p_{n2}, \dots, p_{nN})^T$  is understood:

$$\|p_n\|^2 = \sum_{m=1}^N |p_{nm}|^2 = 1$$

This relation and Equation 7.78 yield

$$P'_s = NP_s$$

For a while, one of  $N$  signals  $x_1$ , for example, is assumed to be sent, that is, others are set to be zero ( $x_n = 0, n = 2, \dots, N$ ). Then Equation 7.75 becomes

$$\begin{aligned} \mathbf{x}' &= (p_{11}, p_{21}, \dots, p_{N1})^T x_1 \\ &= \mathbf{p}_1 x_1 \quad (\mathbf{p}_1 = (p_{11}, p_{21}, \dots, p_{N1})^T) \end{aligned}$$

Received signals  $\mathbf{y} = (y_1, y_2, \dots, y_{N+M})^T$  are given as

$$\mathbf{y} = \mathbf{H}\mathbf{p}_1 x_1$$

Therefore, we have

$$y_n = x_1 \sum_{m=1}^N h_{nm} p_{m1} \quad (n = 1, 2, \dots, N + M)$$

Signals  $\mathbf{y}' = (y'_1, y'_2, \dots, y'_{N+M})^T$  given by multiplying  $\mathbf{y}$  by  $\mathbf{Q}$  become

$$\mathbf{y}' = \mathbf{Q}\mathbf{H}\mathbf{p}_1 x_1 \quad (7.79)$$

or

$$\begin{aligned} y'_i &= \sum_{n=1}^{M+N} q_{in} y_n \\ &= x_1 \sum_{n=1}^{M+N} q_{in} \sum_{m=1}^N h_{nm} p_{m1} \quad (i = 1, 2, \dots, N) \end{aligned}$$

Using the Schwarz inequality,  $|y'_i|^2$  is maximized with

$$q_{in} = k \sum_{m=1}^N h_{nm}^* p_{m1}^* \quad (k : \text{constant}), \quad (i = 1, 2, \dots, N; m = 1, 2, \dots, M + N) \quad (7.80)$$

The right-hand side of the above equation never depends on  $i$ , and we use a notation as

$$\mathbf{q}_1 = (q_{11}, q_{12}, \dots, q_{1N+M})^T$$

where suffix 1 is used to correspond to the vector  $\mathbf{p}_1$ . Then Equation 7.80 is expressed as

$$\mathbf{q}_1 = k \mathbf{H}^* \mathbf{p}_1^*$$

where  $k$  is an arbitrary constant. Hereafter, we put  $k=1$  for simplicity. Similar to the above discussion, for a case where only signal  $x_n$  is not zero, we have

$$\mathbf{q}_n = \mathbf{H}^* \mathbf{p}_n^*$$

Using column vectors  $\mathbf{q}_n$ , we have a matrix as

$$\mathbf{Q}^T = (\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N) = \mathbf{H}^* (\mathbf{p}_1^*, \mathbf{p}_2^*, \dots, \mathbf{p}_N^*) = \mathbf{H}^* \mathbf{P}^*$$

where  $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N)$

Then we get

$$\mathbf{Q} = (\mathbf{H}\mathbf{P})^H = \mathbf{P}^H \mathbf{H}^H \quad (\mathbf{H} : \text{complex conjugate transpose}) \quad (7.81)$$

For the case where only signal  $x_n$  is not zero, using the above expression and Equation 7.79, we have

$$\mathbf{y}' = x_n \mathbf{p}_n^H \mathbf{H}^H \mathbf{H} \mathbf{p}_n$$

From the Schwarz inequality, to maximize  $\|\mathbf{y}'\|$ , we have

$$\mathbf{p}_n = k'_n \mathbf{H}^H \mathbf{H} \mathbf{p}_n$$

Rewriting the above equation, we get

$$\mathbf{H}^H \mathbf{H} \mathbf{p}_n = \lambda_n \mathbf{p}_n \quad (\lambda_n = 1/k'_n) \quad (7.82)$$

The vector  $\mathbf{p}_n$  and a scalar  $\lambda_n$  are, respectively, the *eigenvector* and the *eigenvalue* of the matrix  $\mathbf{H}^H \mathbf{H}$ . Since  $\mathbf{H}^H \mathbf{H}$  is a normal matrix, the eigenvectors are orthogonal with each other, that is,

$$\mathbf{p}_n^H \mathbf{p}_m = \begin{cases} 1 & (m=n) \\ 0 & (m \neq n) \end{cases} \quad (\|\mathbf{p}_n\|^2 = 1)$$

Multiplying both sides of Equation 7.82 by  $\mathbf{p}_n^H$  from the left side yields  $\mathbf{p}_n^H \mathbf{H}^H \mathbf{H} \mathbf{p}_n = \|\mathbf{H} \mathbf{p}_n\|^2 = \lambda_n \|\mathbf{p}_n\|^2$ . Thus, we see the eigenvalues  $\lambda_n$  are positive real numbers.

The equalizer output signal  $\mathbf{y}'$  becomes  $(\|\mathbf{p}_n\|^2 = 1)$ :

$$\mathbf{y}' = x_n \mathbf{p}_n^H \cdot \lambda_n \mathbf{p}_n = \lambda_n x_n \quad (7.83)$$

Since the eigenvalue is a positive real number, polarity of the transmit signal never changed at the output of the receiver.

Now, we send all signals at the same time. Then we have

$$\mathbf{P}\mathbf{x} = \mathbf{x}' = x_1\mathbf{p}_1 + x_2\mathbf{p}_2 + \cdots + x_N\mathbf{p}_N$$

Multiplying the above equation by  $\mathbf{H}^H\mathbf{H}$  from the left side, we get

$$\begin{aligned}\mathbf{H}^H\mathbf{H}\mathbf{P}\mathbf{x} &= \mathbf{H}^H\mathbf{H}\mathbf{x}' = x_1\mathbf{H}^H\mathbf{H}\mathbf{p}_1 + x_2\mathbf{H}^H\mathbf{H}\mathbf{p}_2 + \cdots + x_N\mathbf{H}^H\mathbf{H}\mathbf{p}_N \\ &= \lambda_1x_1\mathbf{p}_1 + \lambda_2x_2\mathbf{p}_2 + \cdots + \lambda_Nx_N\mathbf{p}_N\end{aligned}$$

Multiplying the above equation by  $\mathbf{p}_n^H$  from the left side gives  $\mathbf{p}_n^H\mathbf{H}^H\mathbf{H}\mathbf{P}\mathbf{x} = \lambda_nx_n$ . Using this, we have

$$\begin{aligned}\mathbf{y} &= \mathbf{P}^H\mathbf{H}^H\mathbf{H}\mathbf{P}\mathbf{x} = (\lambda_1x_1, \lambda_2x_2, \cdots, \lambda_Nx_N)^T \\ &= \Lambda\mathbf{x}\end{aligned}\tag{7.84}$$

where  $\Delta(N \times N)$  is a triangular matrix with triangular elements  $\lambda_n$  as

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_N \end{pmatrix}$$

Thus, transmit signals  $x_1 \sim x_N$  are obtained without intersymbol interference.

From Equation 7.84, we have  $\mathbf{P}^H\mathbf{H}^H\mathbf{H}\mathbf{P} = \Lambda$ . Since  $\mathbf{P}$  is a unitary matrix, we have  $\mathbf{P}^H = \mathbf{P}^{-1}$ . Therefore,  $\mathbf{P}^H\mathbf{H}^H\mathbf{H}\mathbf{P} = \mathbf{P}^{-1}\mathbf{H}^H\mathbf{H}\mathbf{P}$  is given. This expression means that our method is based on diagonalization of the matrix  $\mathbf{H}^H\mathbf{H}$  with its eigenvectors.

From Equation 7.84, average receive power becomes

$$\begin{aligned}S_n &= \left\langle |\lambda_nx_n|^2 \right\rangle = \lambda_n^2 \left\langle x_n^2 \right\rangle \\ &= \lambda_n^2 P_s\end{aligned}$$

Denoting noise components after equalization by  $n_n'$ , we have

$$n_n' = \sum_{m=1}^{N+M} q_{nm}z_m$$

The average noise power becomes

$$\begin{aligned}N_n &= \left\langle |n_n'|^2 \right\rangle = \left\langle \left( \sum_{m=1}^{N+M} q_{nm}z_m \right) \left( \sum_{m=1}^{N+M} q_{nm}^*z_m^* \right) \right\rangle \\ &= \sum_{m=1}^{N+M} |q_{nm}|^2 \left\langle |z_m|^2 \right\rangle \\ &= \sum_{m=1}^{N+M} |q_{nm}|^2 \sigma^2 = \|\mathbf{q}_n\|^2 \sigma^2\end{aligned}$$

From Equation 7.81, we have  $\mathbf{q}_n = \mathbf{p}_n^H \mathbf{H}^H$  and

$$\begin{aligned} \|\mathbf{q}_n\|^2 &= \mathbf{p}_n^H \mathbf{H}^H \mathbf{H} \mathbf{p}_n \\ &= \mathbf{p}_n^H \lambda_n \mathbf{p}_n = \lambda_n \quad (\|\mathbf{p}_n\|^2 = 1) \end{aligned}$$

Thus, we get  $N_n = \lambda_n \sigma^2$ .

The signal to noise power ratio is given as

$$\frac{S_n}{N_n} = \lambda_n \frac{P_s}{\sigma^2}$$

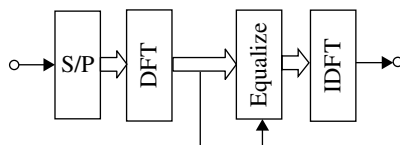
*Discussion:* The above arguments never assume any specific propagation matrix. Furthermore, a signal block (vector) is generally assumed, and it need not to be made up with a single carrier signal in the time domain. Therefore, the above arguments are applicable to any transmission system such OFDM, SDM, and MIMO. The difference in performance and structures of the systems appears only on the propagation matrix and therefore its eigenvectors and eigenvalues.

### 7.3.7 Frequency-Domain Equalizer

As transmission speed and symbol frequency become high, the number of time-resolvable paths in a channel increases. Accordingly, the required number of taps in an equalizer becomes large, and therefore, the calculation complexity increases. To cope with this difficulty in equalization, frequency-domain equalizers (Fig. 7.29) are known [33].

In this system, a received signal block is transformed into the frequency domain with the fast Fourier transform (FFT). Then equalization is carried out in the frequency domain with an estimated frequency transfer function of the channel. The equalized signal is transformed into the time domain with the inverse FFT and then applied to a demodulator.

Reduction of calculation complexity in this system owes to the FFT. If the minimum mean square criterion is used for equalization, the implicit diversity effects (Section 7.1) are obtained. Combining the frequency domain equalization and the decision feedback equalization yields better performance. In this case, the feed-forward part is implemented in the frequency-domain. Although the feedback part is implemented in the time domain, the calculation complexity never becomes a critical issue, since the decided signals are expressed with integer numbers and the number of taps of the feedback part is much less than that of the feed-forward part.



**FIGURE 7.29** Frequency-domain equalization.



For convenience of timing synchronization, a guard period is placed between the neighboring signal blocks, and the tail part of the signal block is copied and inserted in the guard period.

The channel characteristics are estimated as a frequency transfer function by transmitting a reference signal, and the received reference signal is transformed into the frequency domain and compared with that of the transmit reference signal.

### 7.3.8 Turbo Equalizer

Intersymbol interference occurs by addition of prior pulse signals to later pulse signals. The interference signals are produced by convolution of a transmit pulse signal with impulse responses of a channel and a receive filter. A convolutional coding is also carried out by taking convolution of a transmit data signal with an encoder impulse response: a weighted data signal is added to later data signals. The intersymbol interference and the convolutional encoding are equivalent in this sense, although the former is carried out unintentionally and the latter intentionally. Therefore, we can see that replacing a convolutional encoder of a turbo encoder with intersymbol interference (convolution) at a transmission channel shows the effects of the turbo decoding at a receiver. This receiver is named as a turbo equalizer in the paper [34] where the receiver was first proposed.

If the channel impulse response is limited in a finite time period, the state of the received signal is represented with a state transition figure with a finite state. The a priori probability of a data calculated in the equalizer is exchanged as extrinsic information with a decoder as in a turbo receiver. It is shown that the error rate performance is improved as if no intersymbol interference exists.

The calculation complexity of the turbo equalizer increases as  $L^M$  for the impulse response length  $M$  and the number of levels  $L$  of digital signals. To reduce the complexity, approximate methods [35, 36] are known to calculate the a priori probabilities using conventional linear equalizer or decision feedback equalizer.

Estimation accuracy of channel characteristics can be improved iteratively by combining the channel estimation and the turbo decoding [37]. Intersymbol interference is suppressed with channel estimation and the turbo decoding results enjoying a high error correcting capability of the turbo decoding. Monitoring the operation of a receiver such as carrier slip (Section 5.5.1) error becomes sure by using reliabilities of error correction of the turbo decoder.

### 7.3.9 Discussions on Equalization

Noise and intersymbol (signal) interference principally deteriorate receiver performance. In a digital transmission system, noise and the interference at sampling instants become important, since signal decision is made at the sampling instants. Noise power is controlled with only a receive (noise-limiting) filter. The intersymbol interference is, however, affected by transmit filter or transmit pulse waveform, receive filter, and channel distortion. Under a condition for a distortion-free channel and that both transmit and receive filters are controllable, the optimum receiver that shows the

minimum error rates is obtained with the transmit and receive filters that together meet the interference-free (the Nyquist I) condition, and the receive filter is the matched filter (Section 3.3.6). These filters are called the root Nyquist filter and exclusively used for digital radio communication systems. For a channel equalizer, the root Nyquist filters are treated independently; their use is assumed in the arguments so far.

For a distortion channel, error rate performance is improved with signal processing to equalize the channel distortion. If only distortion is considered, an inverse circuit of the channel is applicable as an equalizer at the transmit or receive side. However, if the equalizer is employed at the receive side, noise power increases, and if it is employed at the transmit side, received signal power decreases since the transmit power is assigned more for frequency components, which shows higher attenuation. Therefore, distortion and noise should be considered at the same time for designing an equalizer.

It is difficult to make a strict design criterion for minimization of error rates, since the error rates are highly nonlinear function of noise and interference. The MMSE criterion is usually adopted, where the square value of decision errors at sampling instants is minimized. Under the criterion, an equalizer placed at transmit and/or receive side is designed: the controllable variables of the equalizer are determined to make the mean square errors.

Equalizers are grouped into a linear or nonlinear type. The nonlinear equalizer shows better performance than a linear one, since it uses the property that digital signals take a discrete level.

The MLSE is optimum in a sense that gives mostly likely signals under a distorted channel (Section 3.3.7). This receiver falls in to a nonlinear equalizer, where nonlinear signal processing is done without direct reducing of intersymbol interference. This receiver is possible owing to the fact that a digital signal takes a finite discrete level.

The eigenvector method (vector coding) is based in principle on diagonalization of a channel matrix by linear signal processing at both transmit and receive sides. Intersignal interference is removed with the diagonalization. Therefore, the design criterion becomes to maximize signal to noise power ratio at sampling instants: the receiver becomes a matched filter, and its principle is based on diagonalization of correlation matrix of channel characteristics. The eigenvector method deviates from sending digital signals successively in time; the signals are sent in parallel with multiple long codes (pulses), which are orthogonal with each other. Thus, a data signal assigned a code (eigenvector) with a low eigenvalue shows a low signal to noise power ratio and poor error rate performance. To mitigate this, a higher power should be assigned to the code. From these points, it is close to an OFDM system. The OFDM system also transmits data signals in parallel using orthogonal codes: the cyclic prefix insertion at a guard time period is seen as equalization, where channel characteristics are not fully known and intersymbol interference is not completely removed for a channel with a time dispersion longer than the guard time.

Equalization methods are grouped in the other manner, that is, into those in time domain or frequency domain. The frequency-domain equalizer shows a longer time delay in signal detection due to its block-wise signal processing. However, it has an advantage over a time-domain equalizer: the signal processing complexity becomes less than that for a time-domain equalizer when channel time dispersion becomes

long and the number of the taps of a time-domain equalizer increases. For a time-domain equalizer, its performance can be improved by iterative equalization for a received and stored signal block [38].

A theoretical analysis of error rates for a receiver with an equalizer becomes difficult especially for a nonlinear receiver. An approximate method for the lowest error rates uses the signal to noise power ratio with a matched filter and neglecting the intersymbol interference. The obtained minimum error rates give a reference in performance for various equalizing methods.

### 7.3.10 Applications to a Mobile Radio Channel

From field experiments [39], the highest data rate in a land mobile radio channel without an equalizer was measured to be less than 100 kbps. Adaptive equalizers for application to mobile radio channels were studied corresponding to a proposal for a TDMA mobile telephone system with transmission bit rate of several hundred kbps.

Decision feedback equalizers were studied by Raith, Stjernvall, and Uddenfeldt [40] in frequency-selective fading channels by computer simulation. It was shown that a transmission rate of 300 kbps was feasible with use of an equalizer. It is claimed that the implicit diversity effect obtained by the equalizer as well as some other factors proves advantageous for a TDMA system over a FDMA system. Field tests with the equalizers at transmission rates of 170 and 340 kbps were successful [41]. A decision feedback equalizer with the Kalman algorithm for fast adaptation is described in [42]. Decision feedback equalization together with adjacent channel interference cancellation is discussed in [43]. To cope with both the minimum and nonminimum phase shift conditions, a decision feedback equalizer that has a time-reversal function has been proposed in [44] and [45].

A receiver for digital PM was proposed to use MLSE [46]; thus, it can be applied to a frequency-selective fading channel. The performance of a MLSE Viterbi receiver for fading mobile channels was investigated in [4, 47–52]. For the radio signals for the pan-European digital mobile telephone system, it was shown that the Viterbi equalizer works well for a multipath channel with echo delays on the order of 20  $\mu$ s and vehicle speeds of 200 km/h [50].

Frequency offset is a practical problem for receivers. Frequency offset compensation techniques for the MLSE receiver are described in [50, 51]. Another practical problem with the MLSE receiver is sample timing control. For this purpose, a technique based on double sampling was proposed [4]. The speed of adaptation during a data transmission period must be fast, especially for a long burst length. This is the case for the signals in the US and Japanese digital cellular systems (Section 9.7). The MLSE receiver can cope with a maximum Doppler frequency of 80 Hz [4, 52] in digital cellular systems.

Most equalizers assume a linear channel: the transmitted signal, which may be linearly modulated or not, must be passed through a linear channel. Some channels are not linear, such as an amplitude limiter, a differential detector, or a frequency discriminator. An equalizer for a frequency selective fading channel with a limiter-differential detection system has been proposed [53]. The principle of this method is

based on the learning of intersymbol interference with a training sequence. Corresponding to past decided data and input signals, intersymbol interference is estimated and subtracted from the signal to be decided.

7.4 ERROR CONTROL TECHNIQUES

To cope with errors in digital transmission, we have two countermeasures, namely, error correction and retransmission. Both methods transmit redundant (check or parity) bits along with information bits. The error correction is completed at the receiver side and therefore called forward error correction (FEC). In contrast to the FEC, the retransmission needs information about error from the receiver (feedback request (ARQ). The FEC may fail to correct the error(s) in a bad channel condition. However, the ARQ with a sufficient number of parity bits scarcely fails to detect the errors even in a heavily bad channel condition. Using these differences in performance, both methods are sometimes combined together especially in a data transmission system, and the method is called hybrid (H-)ARQ.

Transmission errors occur due to noise and distortion. We have two types of errors: random errors and burst errors. Random errors may be caused by thermal noise. Burst errors are generated during a fade in the transmission channel. Depending on the type of error, different approaches become useful, as shown later.

Interleaving is a technique for changing burst errors into random error. In this method, a block of data is stored in a two-dimensional table as shown in Figure 7.30. At the transmitter, data is written in the horizontal direction and read and transmitted in the vertical direction. At the receiver, data is written and read in the opposite manner. The burst errors are then separated from each other by the range of the horizontal length of the table. The delay in signal transmission is the penalty with this method. This delay may not be tolerable when interleaving is performed to cope with long

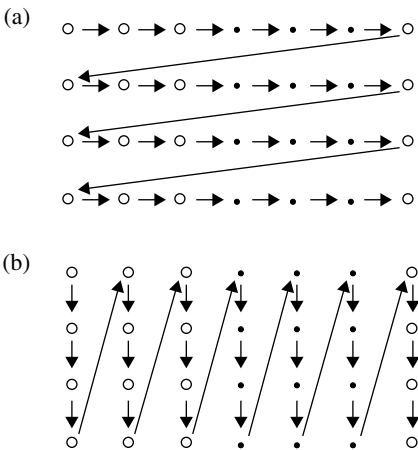


FIGURE 7.30 Bit interleaving. (a) Write and (b) Read.

burst errors caused in a very slowly fading channel. In these circumstances, a diversity transmission or frequency hopping system becomes useful to reduce the length of burst errors.

In this section, a brief description of error control techniques is given. The reader can refer to other books [54, 55] for further discussion.

The most primitive code used for error detection is a parity check code. A check bit is added so that the modulo 2 addition of the coded signal bits is zero (Fig. 7.31). If we denote the information bits as  $(a_1, a_2, \dots, a_n)$ , then the check bit  $c$  is given by  $\sum_{i=1}^n a_i \pmod{2}$ . We can detect a single error or an odd number of errors by checking that  $\sum_{i=1}^{n+1} b_i$  equals 0 or not for the received bits  $b_i$ , that is, the parity of the received bits including the check bit. We can apply parity check codes to the horizontal and vertical directions of a two-dimensional array of data signals (Fig. 7.32). In this case, a single error can be detected and corrected.

A primitive 3-bit code is given by (000) and (111) for data of 0 and 1, respectively. This code is geographically shown in Figure 7.33. One-bit error is corrected or up to two-bit errors are detected.

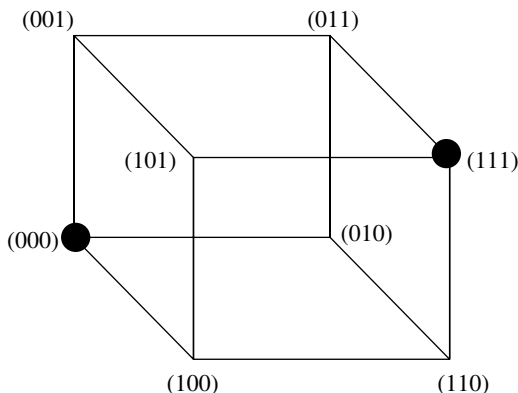
By using more check bits with sophisticated coding algorithms, better performance from error correction and/or detection methods can be achieved. Error correcting codes are classified into two important groups: block codes and convolutional codes. In block codes, a block of  $k$  data bits is encoded to produce a code word of  $n(>k)$  bits. In convolutional codes, a stream of data bits is encoded successively by modulo 2

$a_1$	$a_2$	$a_3$	$c$
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

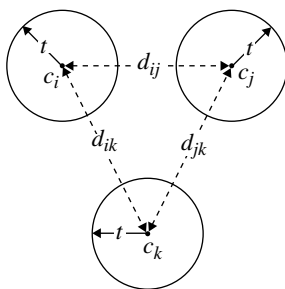
FIGURE 7.31 Parity check coding.

$a_{11}$	$a_{12}$	$a_{13}$	$c_1$
$a_{21}$	$a_{22}$	$a_{23}$	$c_2$
$a_{31}$	$a_{32}$	$a_{33}$	$c_3$
$b_1$	$b_2$	$b_3$	$c_4$

FIGURE 7.32 Horizontal and vertical parity check code.



**FIGURE 7.33** Three-dimensional illustration of a three-bit code.



**FIGURE 7.34** Code words with the minimum distance of  $2t+1$ .

convolutions: a data bit affects the encoded code word for a length of  $N$  bits, called the constraint length (e.g., see Figure 7.36,  $N=3$  in this case).

An important category of error correcting codes is linear block codes. A code word  $\mathbf{c}$  of length  $n$  bits can be expressed as an  $n$ -dimensional vector, whose components are 1 or 0;  $\mathbf{c}=(c_1, c_2, \dots, c_n)$ . For a data word  $\mathbf{d}$  of length  $k$ , we denote  $\mathbf{d}=(d_1, d_2, \dots, d_k)$ . The number of check bits is  $m=n-k$ . The ratio  $k/n$  is known as the code rate or the code efficiency.

In an  $n$ -dimensional space, code words may take  $2^k$  different points, which are a subset of  $2^n$  points in the  $n$ -dimensional space. Error correcting codes are designed so that the minimum number of difference bits between any two code words, called the Hamming distance, is greater than a given value (Fig. 7.34). If the minimum distance is  $2t+1$ , then  $t$ -errors can be corrected or  $2t$  errors can be detected.

A family of codes, where a code word consists of data bits and check bits produced by a linear combination of the data bits, is known as a systematic code.

#### 7.4.1 Linear Block Codes

*From Modern Digital and Analogue Communication systems, Second Edition, by B. P. Lathi. Used by permission from Oxford University Press, Inc [56].*

A linear (systematic) block code is given as

$$\begin{aligned}
 c_1 &= d_1, \\
 c_2 &= d_2, \\
 &\dots, \\
 c_k &= d_k, \\
 c_{k+1} &= h_{11}d_1 \oplus h_{12}d_2 \dots \oplus h_{1k}d_k, \\
 c_{k+2} &= h_{21}d_1 \oplus h_{22}d_2 \dots \oplus h_{2k}d_k, \\
 &\dots, \\
 c_n &= h_{m1}d_1 \oplus h_{m2}d_2 \dots \oplus h_{mk}d_k
 \end{aligned} \tag{7.85}$$

where  $\oplus$  denotes modulo 2 addition. Using matrix expression,

$$\mathbf{c} = \mathbf{d}G$$

where

$$\begin{aligned}
 G &= \begin{bmatrix} 100 & \dots & 0 & h_{11}h_{21} & \dots & h_{m1} \\ 010 & \dots & 0 & h_{12}h_{22} & \dots & h_{m1} \\ 000\dots & \dots & 1 & h_{1k}h_{2k} & \dots & h_{mk} \end{bmatrix} \\
 &= [I_k (k \times k), P(k \times m)]
 \end{aligned}$$

The matrix  $G$  is called the generator matrix. The code vector is expressed as

$$\mathbf{c} = [\mathbf{d}, \mathbf{c}_p]$$

where  $\mathbf{c}_p = \mathbf{d}P$  is a parity check vector.

**Decoding** A received word  $\mathbf{r}$  is given as

$$\mathbf{r} = \mathbf{c} \oplus \mathbf{e}$$

We define a matrix  $H^T$  as

$$H^T = \begin{bmatrix} P \\ I_m \end{bmatrix}$$

where  $I_m$  is the identity matrix of the order  $m \times m$  ( $m = n - k$ ).

The matrix  $H^T$  has the property

$$\begin{aligned}
 \mathbf{c}H^T &= [\mathbf{d}, \mathbf{c}_p] \begin{bmatrix} P \\ I_m \end{bmatrix} \\
 &= \mathbf{d}P \oplus \mathbf{c}_p \\
 &= \mathbf{c}_p \oplus \mathbf{c}_p = \mathbf{0}
 \end{aligned}$$

The transpose of the matrix  $H^T$ ,  $H=[P^T, I_m]$  is called the parity check matrix. We can calculate a vector, called the syndrome, as

$$\begin{aligned} \mathbf{s} &= \mathbf{r}H^T \\ &= (\mathbf{c} \oplus \mathbf{e})H^T \\ &= \mathbf{c}H \oplus \mathbf{e}H^T \\ &= \mathbf{e}H^T \end{aligned}$$

The vectors  $\mathbf{s}$  have  $m(=n-k)$  dimensions;  $\mathbf{s}$  is zero if there is no error ( $\mathbf{e}=\mathbf{0}$ ) in the received vector  $\mathbf{r}$ . Thus, we can decide that there are errors if  $\mathbf{s} \neq \mathbf{0}$ . After calculating syndrome  $\mathbf{s}$ , we proceed to the next step of finding the error vector  $\mathbf{e}$ . If  $\mathbf{e}$  is known, we have the corrected word as  $\mathbf{c} = \mathbf{r} \oplus \mathbf{e}$ . However, the error vector cannot be uniquely determined from the syndrome  $\mathbf{s}$ . There may be multiple combinations of error vectors and data vectors producing the same syndrome. The best strategy, in the sense that the error probability for the corrected word is minimum, is to choose the error vector with the minimum number of 1s as elements (minimum-weight vector). For this error correction method, we can prepare a table of  $2^m$  pairs of minimum-weight error vectors and their corresponding syndromes.

#### 7.4.2 Cyclic Codes

Cyclic codes [56] are a subclass of linear block codes. In cyclic codes a shift of a code becomes another code. From this property of cyclic codes, treatment becomes simple: encoding/decoding is performed with a shift register and a mathematical expression in polynomial form can be used.

The code word with  $n$  bits is expressed as

$$c(x) = c_1x^{n-1} + c_2x^{n-2} + \dots + c_n$$

The coefficients express the elements of the code and take either 0 or 1. Addition of the coefficients is modulo 2 and multiplication obeys the usual integer multiplication. We denote the data polynomial and the generator polynomial by  $d(x)$  and  $g(x)$ , respectively. The systematic cyclic codes are generated as

$$c(x) = x^{n-k}d(x) + \rho(x)$$

where  $\rho(x)$  is the remainder from dividing  $x^{n-k}d(x)$  by  $g(x)$ , that is,

$$\rho(x) = \text{Rem} \frac{x^{n-k}d(x)}{g(x)}$$

and  $g(x) = x^m + g_1x^{m-1} + g_2x^{m-2} + \dots + g_{m-1}x + 1$  ( $m = n - k$ )

Since  $f(x)+f(x)=0$ , the code word  $c(x)$  is a multiple of the generator polynomial  $g(x)$ . The first  $k$  digits of the code word are data digits and the last  $n-k = m$  digits are parity digits.



**Example** Let  $g(x) = x^3 + x + 1$  and  $d(x) = x^3$  (1000); then  $(x^3d(x))/g(x)$  is calculated as

$$\begin{array}{r}
 x^3 + x + 1 \\
 x^3 + x + 1 \sqrt{x^6} \\
 \underline{x^6 + x^4 + x^3} \\
 x^4 + x^2 + x \\
 \underline{x^3 + x^2 + 1} \\
 x^3 + x + 1 \\
 \underline{x^2 + 1}
 \end{array}$$

We have a remainder  $\rho(x) = x^2 + x$  and the encoded word  $c(x) = x^6 + x^2 + 1$  or (1000101).

In usual manner of division of polynomial (without using mod 2 addition or subtraction), the above encoding process is understood as follows.

We can see that the remainder with the usual manner has negative coefficients as

$$\rho'(x) = \text{Rem} \frac{x^{n-k}d(x)}{g(x)} = -\rho(x)$$

since  $x^m d(x)$  has zero coefficients for  $x^n (n < m)$ . Then subtraction of  $\rho'(x)$  from  $x^m d(x)$  becomes addition of  $\rho(x)$ .

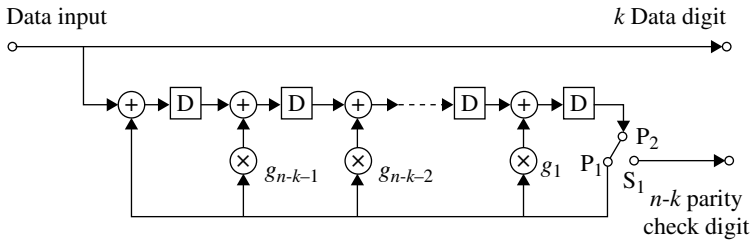
Furthermore, if we do not use the polynomial expression but use the usual integer number expression for the data encoding, the systematic cyclic encoding process is understood as follows. The term  $x^m d(x)$  corresponds to a number  $X_d = (d_1, d_2, d_3, \dots, d_k, 0, 0, \dots, 0)$  where 0 continues  $m$ -times. The generator polynomial corresponds to  $G = (g_1, g_2, g_3, \dots, g_{m+1})$ . Then we have

$$\text{Rem} \frac{X_d}{G} = R (\geq 0)$$

If we subtract  $R$  from  $X_d$ , no systematic codes are given. However, if we add  $(0 \leq) R' (= G - R) (\leq G)$ , we have a systematic code, where the number  $G - R$  corresponds to  $\rho(x)$ .

Denote the error polynomial by  $e(x)$ ; then the syndrome polynomial  $s(x)$  becomes

$$\begin{aligned}
 s(x) &= \text{Rem} \frac{r(x)}{g(x)} \\
 &= \text{Rem} \frac{c(x) + e(x)}{g(x)} \\
 &= \text{Rem} \frac{e(x)}{g(x)}
 \end{aligned}$$



**FIGURE 7.35** Encoder for cyclic codes.

If there is no error or a code word is received that may be different from the transmitted one, we have  $s(x)=0$ . For deciding on an error polynomial  $e(x)$  from  $s(x)$ , we choose the one (the minimum-weight vector) that has the minimum number of 1's as its coefficients.

A cyclic code can be generated by a feedback shift register as shown in Figure 7.35. The switch  $S_1$  is at the positions  $P_1$ .  $k$  data bits followed by  $n-k$  successive 0's are fed into the shift register. As the  $k$ th data bit clears out of the last register (the content of the registers at this moment corresponds to the remainder), the switch is thrown to the position  $P_2$  and the  $m(=n-k)$  parity check bits are output. The feedback shift register performs the division  $d(x)x^{n-k}/g(x)$ . Hence, the syndrome  $s(x)$  can be obtained by the same feedback shift register.

The ability to correct/detect errors is given by the generator polynomial. The code with a minimum distance  $2t+1$  can detect up to  $2t$  errors. A discussion of the burst error-detecting capability of a cyclic code follows. A burst error of length  $L$  bits is expressed as

$$e(x) = x^i b(x) \quad (i = 0, 1, 2, \dots, n-L)$$

where  $b(x)$  is a polynomial expressing the error pattern:

$$b(x) = x^{L-1} + b_2 x^{L-2} + \dots + 1$$

Assume a generator polynomial  $g(x)$  of degree  $m = n-k$ . We can detect any burst error of length  $L \leq m$ , since  $e(x)$  is not a factor of  $g(x)$  in this case. A burst error of length  $L > m$  can be detected with some probability. We have  $2^{L-2}$  error patterns. The number of error patterns that may be a factor of  $g(x)$  is 1 for  $L = m+1$  and  $2^{L-m-2}$  for  $L > m+1$ . The probability that we cannot detect the error becomes

$$P_m = \begin{cases} \frac{1}{2^{m-1}} & (L = m+1) \\ \frac{1}{2^m} & (L > m+1) \end{cases} \quad (7.86)$$

Standardized error-detecting codes known as CRC (cyclic redundancy check) are given by CCITT, one of which is given as

$$\begin{aligned} g(x) &= x^{16} + x^{12} + x^5 + 1 \\ &= (x+1)(x^{15} + x^{14} + x^{13} + x^{12} + x^4 + x^3 + x^2 + 1) \end{aligned}$$

It can be shown that the minimum distance of this code is  $d_{\min}=4$ . It can detect up to 3 errors. Since  $g(x)$  has a factor of  $x+1$  and  $g(-1)=g(1)=c(1)=0$ , any odd number of errors can be detected. Since the degree of  $g(x)$  is 16, we can detect any burst error with length of less than or equal to 16. The probability that a burst error of length greater than 16 cannot be detected is given by Equation 7.86 with  $m=16$ .

### 7.4.3 Convolutional Codes

It is difficult to describe convolutional codes using a mathematical expression; therefore, we discuss some specific examples. A convolutional encoder with a code rate of  $1/2$  is given in Figure 7.36. The encoder outputs two bits for every one input bit. The output bits are determined from the input bit and the two previous input bits stored in the shift register. Thus, this is an 8-state machine. Although it has eight states, its input-output relation is uniquely described by the 4-state diagram shown in Figure 7.37, where state 10, for example, corresponds to the registers  $D_1$  and  $D_2$  containing 1 and 0, respectively. Once an input bit is received, the state changes or

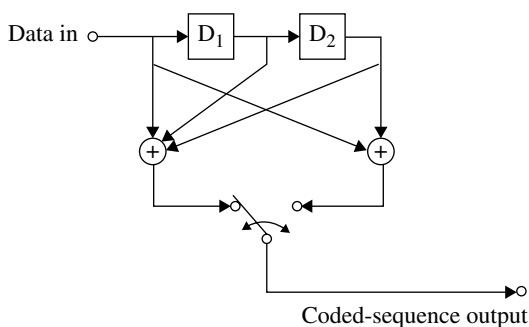


FIGURE 7.36 Encoder for a convolutional code.

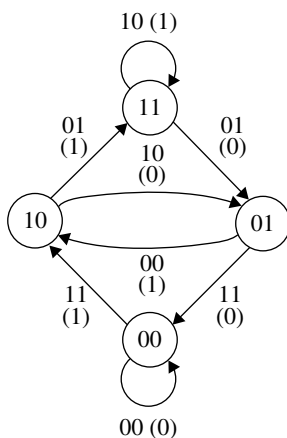


FIGURE 7.37 State diagram of the encoder in Figure 7.36.

remains the same, outputting two bits as shown on the branches. The input bit is shown in parentheses.

Since the input data affects the three successors on code words, decoding with the maximum likelihood sequence estimator or the Viterbi algorithm discussed in Section 3.3.8 is useful.

Figure 7.38a and b shows examples of the decoding process assuming the same transmit data sequence. The Hamming distance is used for the branch metric. In the case of no transmission errors (Fig. 7.38a), path merge, where all the survived paths are originated from a state (01), occurs at time 7. We can decide the data before each merge. Even if the error occurs (Fig. 7.38b), data are correctly decoded. However, the decision on data is delayed since the path merge is delayed due to the error.

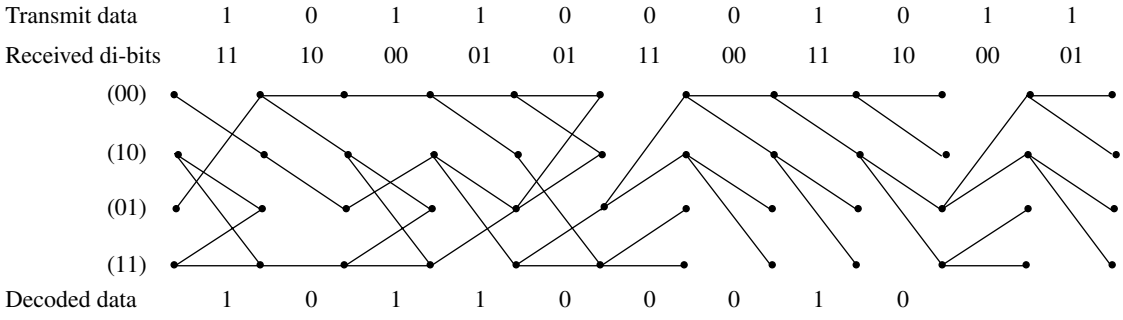
#### 7.4.4 Concatenated Codes

A concatenated code consists of two codes for improvement of error-correcting capability (Fig. 7.39). A code near the channel is called inner code and the other is called outer code. The inner code such as a convolutional code is selected to correct random errors, and the outer code such as the Reed Solomon code is selected to correct burst errors. Such a code assignment is preferable, since the output of the inner code tends to generate a burst error if it fails to correct errors. In some cases an interleaver is inserted between the inner and the outer codes to further improve burst error-correcting capability. The concatenated two code can be thought equivalently as a code with a very long code length resulting into high error-correcting capability. The signal processing complexity never increases regardless of the long code length, since decoding is carried out independently for the inner and the outer code. A concatenated code was used for radio communication between an earth station and a spaceship for deep space investigations. It is applied also for satellite broadcasting and digital video disks. A concatenated code adopts usually a tandem connection of two codes. However, a code with parallel connection such as the turbo codes falls into the category of concatenated codes in a sense that multiple codes are combined together. The concatenated code had shown the highest error-correcting capability till the turbo codes or the low-density parity check (LDPC) codes have appeared.

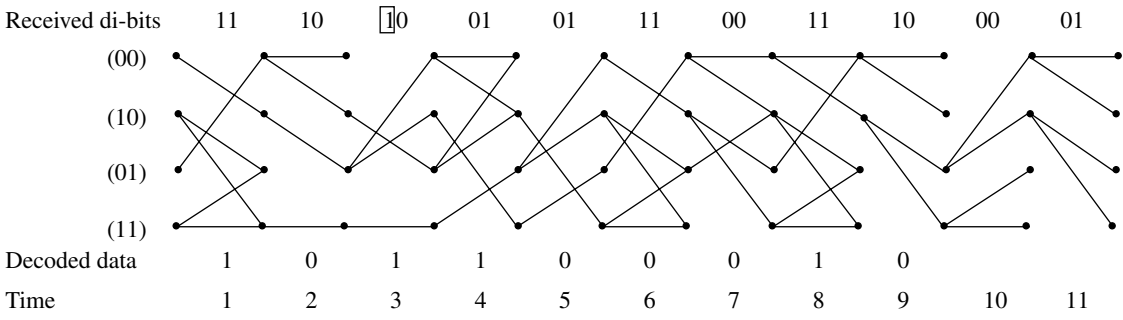
#### 7.4.5 Turbo Codes

This code is proposed by Berrou, Glavieux, and Thitimajashima in 1993 [57, 58]. It is a breakthrough in error correction techniques approaching close to the Shannon limit (Section 3.8) performance. Input data are coded directly with a coder to produce parity bits, and the same input data are time interleaved and coded with another coder (Fig. 7.40) to produce the other parity bits. The two coded data (parity bits) streams are multiplexed with the original input data (systematic codes), and the multiplexed signal is sent into a channel. Two times transmission of the coded parity bits of the input data with and without interleaving increases decoding reliabilities by producing a priori information for the received data. Puncturing the encoded parity bits is usually done to increase its code rate.

(a)



(b)



**FIGURE 7.38** An example of Viterbi decoding process of a convolutional code. (a) Without transmission error. (b) With transmission error. The data in the box is erroneous.

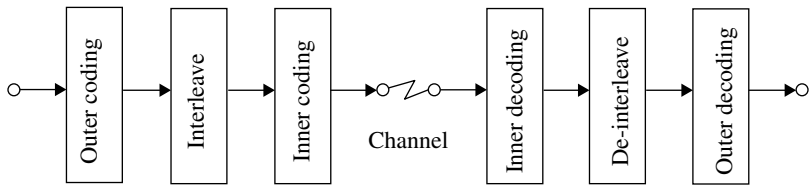


FIGURE 7.39 Concatenated codes.

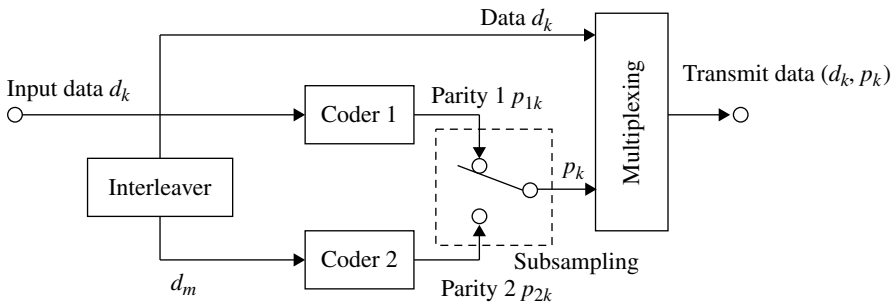


FIGURE 7.40 An example of the systematic turbo code.

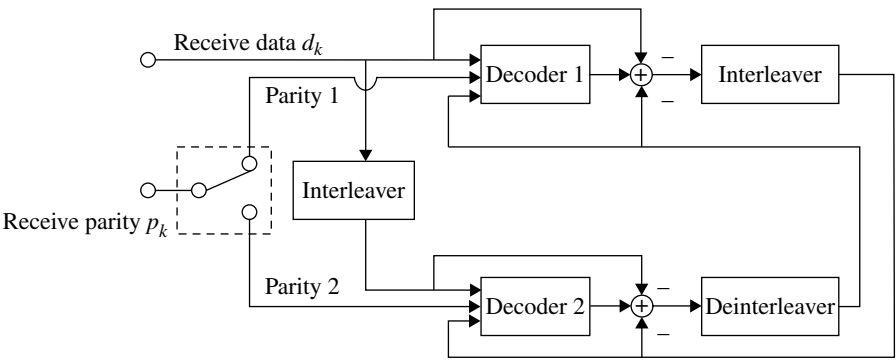
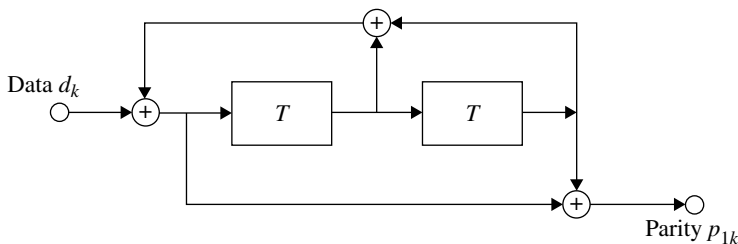


FIGURE 7.41 Turbo code decoder.

Decoding is carried out corresponding to the two encoders (Fig. 7.41). A block of the received data and parity bits is stored for iterative decoding. A coder that corresponds to the encoder calculates reliability of each data by using received soft (analog value) data, its parity bits, and the a priori probabilities (reliability information) that are given by the other decoder. The calculated reliabilities for each data are subtracted by the given a priori probabilities to produce an extrinsic information, and the extrinsic information is sent to the other decoder as the a priori probabilities. The other decoder performs the same procedure. This feedback process is repeated for a given cycle. This code is named as the turbo codes in analogy to the turbo engine where a part of the output power is used to compress the input air to increase the output power. The reliabilities of the a priori probabilities and therefore error rate



**FIGURE 7.42** An example of convolutional code encoder.

performances are improved to saturated values through the iterative decoding exchanging the updated extrinsic information. The bit error rates are minimized with the turbo decoding in contrast to the MLSE.

We proceed to explain the turbo codes using figures and mathematical expressions [59, 60]. Figure 7.42 shows an encoder (encoder 1) for an example. Input data  $d_k$  are applied to a recursive convolutional encoder to produce parity bits  $p_{1k}$ . The data  $d_k$  are interleaved and the output data  $d_n$  are given. The data  $d_n$  are encoded with the other encoder (encoder 2) to produce parity bits  $p_{2k}$ . The encoder 2 may be the same as the encoder 1. When code rate is intended to increase, the parity bits are subsampled (punctured) (Fig. 7.40). The coding rate becomes 1/2 for the turbo coder in Figure 7.40. The feedback convolutional coder in Figure 7.42 has two-stage memory circuits. A parity bit is generated by modulo 2 addition of input data  $d_k$  and the outputs of the two memory circuits. The four possible combinations of the contents  $\{1, 0\}$  of the memory circuits are denoted as (0,0), (0,1), (1,0), (1,1) corresponding to the contents of the first and second memory circuits, respectively. Alternatively, these states are described as  $m=0,1,2,3$  in the binary expression. The state transitions are shown in Figure 7.43, where  $s_k$  denotes a state at time  $k$  and the initial state of  $m=0$  is assumed. The solid and broken lines correspond to input data of 1 and 0, respectively. Data  $d_k$  of ( $d_0 \sim d_5$ ) are assumed for simplicity. The state is forcibly terminated to the initial state of  $m=0$  by adding data  $d_6, d_7$ , which are determined depending on state  $s_6$ . A path corresponds uniquely to a given transmit data sequence  $d_k (k=1,2,\dots,N)$ . Therefore, the data sequence is determined if the path is estimated.

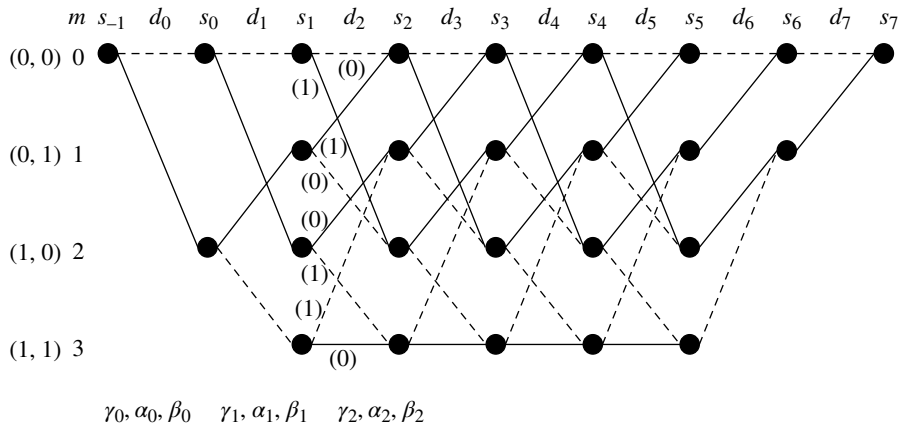
Now, we assume to send a data sequence  $\mathbf{d} = \{d_1, d_2, \dots, d_N\} (d_k = 0 \text{ or } 1)$ . Let us denote the parity sequence generated through turbo coders by  $\mathbf{p} = \{p_1, p_2, \dots, p_N\} (p_k = 0 \text{ or } 1)$ .

If the transmit pulse signal takes a voltage level of  $\pm 1$ , the data signal  $x_k$  and parity signal  $y_k$  are given as

$$\begin{aligned} x_k &= 2d_k - 1 \\ y_k &= 2p_k - 1 \end{aligned}$$

or

$$x_k = \begin{cases} 1 & (d_k = 1) \\ -1 & (d_k = 0) \end{cases}, \quad y_k = \begin{cases} 1 & (p_k = 1) \\ -1 & (p_k = 0) \end{cases}$$



**FIGURE 7.43** State transition graph (solid line for  $d_k = 1$  and broken line for  $d_k = 0$ . Number in parentheses corresponds parity bit).

We assume a static distortion-free channel with additive white Gaussian noise and that  $x_k$  and  $y_k$  are sent with QPSK modulation. Then the received signals corresponding to the data and parity are expressed as

$$\begin{aligned} R_{d_k} &= \rho x_k + n_{d_k} \\ R_{p_k} &= \rho y_k + n_{p_k} \end{aligned}$$

where the suffixes  $d_k$  and  $p_k$  denote the data and parity, respectively, and  $\rho$  the channel gain, and  $n_{d_k}, n_{p_k}$  independent Gaussian noises.

**Log-Likelihood Ratio.** To prepare for description of the turbo decoding, we explain the log-likelihood ratio (LLR). The LLR is defined as logarithm of the ratio of probabilities that a digital signal takes one of the two levels as follows:

$$L(d_k) \equiv \ln \frac{P_r(d_k = 1)}{P_r(d_k = 0)} \left( = \ln \frac{P_r(x_k = 1)}{P_r(x_k = -1)} \equiv L(x_k) \right)$$

If  $P_r(d_k = 1) > P_r(d_k = 0)$ , then  $L(d_k) > 0$ ; otherwise,  $L(d_k) < 0$ . Since  $P_r(d_k = 0) = 1 - P_r(d_k = 1)$ , we have

$$e^{L(d_k)} = \frac{P_r(d_k = 1)}{1 - P_r(d_k = 1)}$$

or

$$\begin{aligned} P_r(d_k = 1) &= \frac{e^{L(d_k)}}{1 + e^{L(d_k)}} \\ &= \frac{1}{1 + e^{-L(d_k)}} \end{aligned}$$



Similarly, we get

$$P_r(d_k = 0) = \frac{e^{-L(d_k)}}{1 + e^{-L(d_k)}}$$

Thus, we have

$$P_r(d_k)(= P_r(x_k)) = \frac{e^{-L(d_k)/2}}{1 + e^{-L(d_k)}} e^{x_k L(d_k)/2} \quad (7.87)$$

The LLR for a received signal  $R_{d_k}$  with transmit data  $d_k$  is defined as

$$L(R_{d_k} | d_k) \equiv \ln \frac{P_r(R_{d_k} | d_k = 1)}{P_r(R_{d_k} | d_k = 0)} \quad (7.88)$$

Under Gaussian noise, we have

$$P_r(R_{d_k} | d_k) = P_r(R_{d_k} | x_k) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(R_{d_k} - \rho x_k)^2}{2\sigma^2}}$$

where  $\rho x_k$  and  $\sigma^2$  are the sampled received signal and noise power at the output of the receive filter at time instant  $k$ .

Thus, we have

$$\begin{aligned} L(R_{d_k} | d_k) &= \frac{2\rho}{\sigma^2} R_{d_k} \\ &= L_c R_{d_k} \quad \left( L_c \equiv \frac{2\rho}{\sigma^2} \right) \end{aligned}$$

Similarly, we have

$$\begin{aligned} P_r(R_{p_k} | p_k) &= P_r(R_{p_k} | y_k) \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(R_{p_k} - \rho y_k)^2}{2\sigma^2}} \end{aligned}$$

Consider that the matched filter is used at a receiver. The transmit pulse waveform and the transmit signal are denoted as  $p(t)$  and  $Ax_k p(t)$ , respectively. A received signal is expressed as

$$r_k(t) = LAx_k p(t) + n(t)$$

where  $L$  denotes the channel gain and  $n(t)$  the noise.

We denote the received signal at the output of the matched filter at a sampling instant  $t=t_m$  by

$$R_{d_k} (= r_k(t) * p(t_m - t) | t = t_m).$$

Letting the signal part by  $\rho x_k$ , we have

$$\rho = LA \int p^2(t) dt,$$

$$\frac{E_b}{N_0} = \frac{\rho^2}{\sigma^2}$$

where  $E_b = (LA)^2 \int p^2(t) dt$  is the energy per bit and  $N_0$  is the double-side-band noise power spectral density of the baseband white Gaussian noise  $n(t)$ . Then we have  $\sigma^2 = N_0 \int p^2(t) dt$ .

Using the above relations, we get

$$P_r(R_{d_k} | x_k) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{E_b}{2N_0}(R_{d_k}/\rho - x_k)^2} \quad (7.89)$$

Then we have

$$L(R_{d_k} | d_k) = 2 \frac{E_b}{N_0} \frac{R_{d_k}}{\rho} \quad (7.90)$$

Similarly,

$$P_r(R_{p_k} | y_k) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{E_b}{2N_0}(R_{p_k}/\rho - y_k)^2} \quad (7.91)$$

$$L(R_{p_k} | p_k) = 2 \frac{E_b}{N_0} \frac{R_{p_k}}{\rho}$$

**The Maximum A Posteriori Algorithm.** We denote a received signal sequence as  $\mathbf{R} = \{R_1, R_2, \dots, R_N\}$  ( $R_k = (R_{d_k}, R_{p_k})$ ).

On receiving  $\mathbf{R}$ , we estimate the most probable transmit data  $d_k$  ( $k = 1, 2, \dots, N$ ). In other words, we decide  $a \in \{1, 0\}$  that maximizes the a posteriori probability of  $d_k = a$  for a given received signal sequence  $\mathbf{R}$ , expressed as

$$P_r(d_k = a | \mathbf{R})$$

The LLR is defined as

$$L(d_k | \mathbf{R}) \equiv \ln \frac{P_r(d_k = 1 | \mathbf{R})}{P_r(d_k = 0 | \mathbf{R})} \quad (7.92)$$

We estimate as  $\hat{d}_k = 1$  if  $L(d_k | \mathbf{R}) > 0$  and  $\hat{d}_k = 0$  if  $L(d_k | \mathbf{R}) < 0$ .

As is seen from the state transition diagram,

$$P_r(d_k = a | \mathbf{R}) = \sum_{\mathbf{d}_i | d_k = a} P_r(\mathbf{d}_i | \mathbf{R})$$

where the summation is made for all the data sequence that takes  $d_k = a$ . A given sequence is denoted as  $\mathbf{d}_i = \{d_{ij}\} (j = 1, 2, \dots, N)$  hereafter. Similarly, a parity sequence is denoted as  $\mathbf{p}_i = \{p_{ij}\} (j = 1, 2, \dots, N)$ . The total number of  $\mathbf{d}_i$  and  $\mathbf{p}_i$  increases exponentially with  $N$ .

Since it is difficult to calculate the a posteriori probabilities, we use the Bayes' theorem as

$$P_r(a, b) = P_r(b | a)P_r(a) = P_r(a | b)P_r(b)$$

Then we have

$$P_r(a | b) = \frac{P_r(b | a)P_r(a)}{P_r(b)}$$

where  $P_r(a, b)$  is the probability that  $a$  and  $b$  occur simultaneously and  $P_r(a | b)$  is the a posteriori probability of  $a$  for a given  $b$ . From the above relation, we get

$$\begin{aligned} P_r(d_k = a | \mathbf{R}) &= \frac{P_r(\mathbf{R}, d_k = a)}{P_r(\mathbf{R})} \\ &= \frac{P_r(\mathbf{R} | d_k = a)P_r(d_k = a)}{P_r(\mathbf{R})} \\ &= \sum_{\mathbf{d}_i | d_k = a} P_r(\mathbf{R} | \mathbf{d}_i)P_r(\mathbf{d}_i) / P_r(\mathbf{R}) \end{aligned} \quad (7.93)$$

If  $d_k$ ,  $n_{dk}$ , and  $n_{pk}$  are independent with each other, we have

$$\begin{aligned} P_r(\mathbf{R} | \mathbf{d}_i)P_r(\mathbf{d}_i) &= \prod_{j=1}^N P_r(R_j | d_{ij})P_r(d_{ij}) \\ &= \prod_{j=1}^N \gamma_{ij} \end{aligned} \quad (7.94)$$

where

$$\gamma_{ij} \equiv P_r(R_j | d_{ij})P_r(d_{ij}) = P_r(R_j, d_{ij})$$

From Equations 7.92–7.94, we get

$$L(d_k | \mathbf{R}) = \ln \frac{\sum_{\mathbf{d}_i | d_k = 1} \prod_{j=1}^N \gamma_{ij}}{\sum_{\mathbf{d}_i | d_k = 0} \prod_{j=1}^N \gamma_{ij}} = \ln \frac{P_{AP}(d_k = 1)}{P_{AP}(d_k = 0)} \quad (7.95)$$

where

$$P_{AP}(d_k = a) \equiv \sum_{\mathbf{d}_i | d_k = a} \prod_{j=1}^N \gamma_{ij} \quad (7.96)$$

A straightforward calculation of the above equation requires  $(N-1)2^N$  times multiplication of  $\gamma_{ij}$ . To reduce the computational complexity, a method [57] that modified the BCJR algorithm [61] is known as follows.

We modify Equation 7.96 as

$$P_{AP}(d_k = a) = \sum_{\mathbf{d}_i | d_k = a} \left( \prod_{j=0}^{k-1} \gamma_{ij} \right) \gamma_{ik} \left( \prod_{j=k+1}^{N-1} \gamma_{ij} \right)$$

For a convolutional code, referring to the state transition diagram (Fig. 7.43), this formula can be rewritten as (instead of making summation after  $N$ -times multiplications, multiplications are divided into smaller times multiplications, summation is made for the divided multiplications, the summation results are multiplied, and the summation of the last results is made; the total number of multiplications is reduced with this modification)

$$\begin{aligned} P_{AP}(d_k = a) &= \sum_{(s' \rightarrow s); d_k = a} \left\{ \left( \sum_{\mathbf{d}_i | s_{k-1} = s'} \prod_{j=1}^{k-1} \gamma_{ij} \right) \gamma_k(s', s) \left( \sum_{\mathbf{d}_i | s_k = s} \prod_{j=k+1}^N \gamma_{ij} \right) \right\} \\ &\equiv \sum_{(s' \rightarrow s); d_k = a} \alpha_{k-1}(s') \gamma_k(s', s) \beta_k(s) \end{aligned} \quad (7.97)$$

where  $\sum_{(s' \rightarrow s); d_k = a}$  means summation over all transitions  $(s_{k-1} = s' \rightarrow s_k = s)$  corresponding to  $d_k = a$  and

$$\alpha_{k-1}(s') = \sum_{\mathbf{d}_i | s_{k-1} = s'} \prod_{j=1}^{k-1} \gamma_{ij}$$

where  $\sum_{\mathbf{d}_i | s_{k-1} = s'}$  means summation over all of sequences  $d_{i1} \sim d_{i(k-1)} (i=1, 2, \dots)$  that take a state  $s_{k-1} = s'$  at time  $k-1$ . Furthermore,

$$\beta_k(s) = \sum_{\mathbf{d}_i | s_k = s} \prod_{j=k+1}^N \gamma_{ij}$$

where  $\sum_{\mathbf{d}_i | s_k = s}$  means summation over all of sequences  $d_{ik} \sim d_{iN} (i=1, 2, \dots)$  originating from a state  $s_k = s$ .

$\gamma_k(s', s)$  is a combined probability of  $R_k$  and  $s_k = s$  under  $s_{k-1} = s'$  and is given as

$$\begin{aligned} \gamma_k(s', s) &= P_r(\{R_k, s_k = s\} | s_{k-1} = s') \\ &= P_r(R_k | \{s_{k-1} = s', s_k = s\}) P_r(s_k = s | s_{k-1} = s') \end{aligned}$$

We will investigate what the equation means in the following.

Consider a pair of states with  $s_{k-1} = s'$  and  $s_k = s$ . If the transition for the pair never exists, we have

$$P_r(s_k = s | s_{k-1} = s') = 0$$

If the transition exists, we get

$$P_r(s_k = s \mid s_{k-1} = s') = P_r(d_k = a)$$

where the value of  $a \in \{1, 0\}$  is given with values of  $(s', s)$ .

If a transition exists,  $d_k$  and  $p_k$  are uniquely given with  $s'$  and  $s$ .

For a given sequence  $\mathbf{d}_i$ ,  $s_k$  and  $d_k$  ( $k=1, 2, \dots, N$ ) are uniquely determined. If we take  $s'$  and  $s$  as those in the sequence, we get

$$\begin{aligned} P_r(R_k \mid \{s_{k-1} = s', s_k = s\}) &= P_r(R_k \mid d_{ik}), \\ P_r(d_k = a) &= P_r(d_{ik} = a) \end{aligned}$$

Then we have

$$\gamma_k(s', s) = \gamma_{ik}$$

Therefore, we obtain

$$L(d_k \mid \mathbf{R}) = \ln \frac{\sum_{\mathbf{d}_i \mid d_k=1} \prod_{j=1}^N \gamma_{ij}}{\sum_{\mathbf{d}_i \mid d_k=0} \prod_{j=1}^N \gamma_{ij}} = \ln \frac{\sum_{\mathbf{d}_i \mid d_k=1} \prod_{k=1}^N \gamma_k(s', s)}{\sum_{\mathbf{d}_i \mid d_k=0} \prod_{k=1}^N \gamma_k(s', s)} \quad (7.98a)$$

where  $s'$  and  $s$  are taken as those determined with  $\mathbf{d}_i$ .

As seen from the state transition diagram,  $\alpha_{k-1}(s')$  and  $\beta_k(s)$  are calculated with a recurrence equation as

$$\alpha_k(s) = \sum_{\text{all } s'} \alpha_{k-1}(s') \gamma_k(s', s) \quad (7.98b)$$

where the summation is taken over the transitions to a state  $s_{k-1} = s$  from all the states  $s'$  of  $s_{k-2}$ .

$\beta_k(s)$  is calculated with a recurrence equation with inverse direction as

$$\beta_{k-1}(s') = \sum_{\text{all } s} \beta_k(s) \gamma_k(s', s)$$

where the summation is taken over the transitions originating from  $s_{k-1} = s'$  to all states  $s$  of  $s_k$ .

The initial value  $\alpha_0(s')$  in the calculation of  $\alpha_{k-1}(s')$  is set as  $\alpha_0(s') = 1$  for  $s' = 0$  and  $\alpha_0(s') = 0$  for  $s' \neq 0$  assuming the initial state of  $s_0 = 0$ .

The initial value  $\beta_N(s)$  in calculation of  $\beta_k(s)$  depends on that the last state is terminated with a fixed value or not. If the state is terminated with  $s_N = 0$ , we use  $\beta_N(s) = 1$  for  $s = 0$  and  $\beta_N(s) = 0$  for  $s \neq 0$ . It is difficult for us to terminate the last state of the two feedback convolutional coders with a fixed value by adding a data through the interleaver. Therefore, actually the terminating data are applied independently to

the two coders [59]. If we do not terminate the state with a fixed value [60], we take either  $\beta_N(s) = 1$  for all  $s$  or

$$\beta_N(s) = \begin{cases} 1 & (s = 0) \\ 0 & (s \neq 0) \end{cases}$$

From the arguments so far, our task reduced to the calculation of  $\gamma_k(s', s) (k = 1, 2, \dots, N)$ . Now, we proceed to the calculation. We restrict state transitions  $s_{k-1} = s' \rightarrow s_k = s$  to the allowed ones. Then we have

$$\gamma_k(s', s) = P_r(R_k | \{s', s\}) P_r(d_k)$$

Since  $\{s', s\}$  is given uniquely with  $\{d_k, p_k\}$ , we get

$$P_r(R_k | \{s', s\}) = P_r(R_k | \{d_k, p_k\})$$

where

$$R_k = (R_{d_k}, R_{p_k})$$

$R_{d_k}$  and  $R_{p_k}$  are independent with each other, and we have

$$P_r(R_k | \{d_k, p_k\}) = P_r(R_{d_k} | d_k) P_r(R_{p_k} | p_k)$$

Therefore, we obtain

$$\gamma_k(s', s) = P_r(R_{d_k} | d_k) P_r(R_{p_k} | p_k) P_r(d_k) \quad (7.99a)$$

If we have multiple parity bits for a data, denoting these by  $p_{kl}$ , we have

$$P_r(R_{p_k} | p_k) = \prod_{l=1}^n P_r(R_{p_{kl}} | p_{kl})$$

If the parity bits are subsampled and  $R_{p_k}$  never corresponds to  $p_k$ , it is set to be a constant or equivalently  $R_{p_k} = 0$  is used in the calculation of  $P_r(R_{p_k} | p_k)$ , since  $P_r(R_{p_k} | p_k)$  never depends on  $p_k = a$ .

In the following,  $\gamma_k(s', s)$  (Eq. 7.99a) is expressed with received signals.

For the first, from Equation 7.87, we have

$$\begin{aligned} P_r(d_k) &= P_r(x_k) \\ &= c_1 e^{x_k L(d_k)/2} \end{aligned} \quad (7.99b)$$

where

$$c_1 = \frac{e^{-L(d_k)/2}}{1 + e^{-L(d_k)}}$$

From Equations 7.89 and 7.91, we get

$$\begin{aligned}
 P_r(R_{d_k} | d_k) P_r(R_{p_k} | p_k) &= P_r(R_{d_k} | x_k) P_r(R_{p_k} | y_k) \\
 &= \frac{1}{2\pi\sigma^2} \exp\left(-\frac{E_b}{2N_0} \left(\frac{R_{d_k}}{\rho} - x_k\right)^2\right) \exp\left(-\frac{E_b}{2N_0} \left(\frac{R_{p_k}}{\rho} - y_k\right)^2\right) \\
 &= \frac{1}{2\pi\sigma^2} \exp\left(-\frac{E_b}{2N_0} \left(\left(\frac{R_{d_k}}{\rho}\right)^2 + \left(\frac{R_{p_k}}{\rho}\right)^2\right.\right. \\
 &\quad \left.\left.+ 2 - 2\frac{R_{d_k}}{\rho} x_k - 2\frac{R_{p_k}}{\rho} y_k\right)\right) \\
 &= c_2 \cdot c_3 \exp\left(\frac{E_b}{N_0} \frac{1}{\rho} (R_{d_k} x_k + R_{p_k} y_k)\right) \tag{7.100}
 \end{aligned}$$

where

$$\begin{aligned}
 c_2 &= \frac{1}{2\pi\sigma^2} \exp\left(-\frac{E_b}{2N_0\rho^2} (R_{d_k}^2 + R_{p_k}^2)\right) \\
 c_3 &= \exp\left(-\frac{E_b}{N_0}\right)
 \end{aligned}$$

Therefore, we have

$$\gamma_k(s', s) = c_1 c_2 c_3 e^{x_k L(d_k)/2} \cdot \exp\left(\frac{E_b}{N_0\rho} (R_{d_k} x_k + R_{p_k} y_k)\right) \tag{7.101}$$

$c_1$ ,  $c_2$ , and  $c_3$  never depend on  $x_k$  and  $y_k$ , and these terms disappear when the ratio of  $\gamma_k(s', s)$  (Eq. 7.98a) is calculated for  $x_k = 1 (d_k = 1)$  and  $x_k = -1 (d_k = 0)$  or  $y_k = \pm 1$ .

**Iterative Turbo Decoding.** Applying Equations 7.97 and 7.99a to Equation 7.95, we have

$$\begin{aligned}
 \Lambda(d_k) &\equiv \ln \frac{P_r(d_k = 1 | \mathbf{R})}{P_r(d_k = 0 | \mathbf{R})} = \ln \frac{P_r(R_{d_k} | d_k = 1) P_r(d_k = 1)}{P_r(R_{d_k} | d_k = 0) P_r(d_k = 0)} \frac{\sum_{(s' \rightarrow s): d_k = 1} \alpha_{k-1}(s') \gamma_k^p(s', s) \beta_k(s)}{\sum_{(s' \rightarrow s): d_k = 0} \alpha_{k-1}(s') \gamma_k^p(s', s) \beta_k(s)} \\
 &\equiv L_{\text{ch}}(d_k) + L_a(d_k) + L_e(d_k)
 \end{aligned}$$

where

$$\begin{aligned}
 L_{\text{ch}}(d_k) (= L(R_{d_k} | d_k)) &= \ln \frac{P_r(R_{d_k} | d_k = 1)}{P_r(R_{d_k} | d_k = 0)}, \\
 L_a(d_k) (= L(d_k)) &= \ln \frac{P_r(d_k = 1)}{P_r(d_k = 0)},
 \end{aligned}$$

and

$$L_e(d_k) = \ln \frac{\sum_{(s' \rightarrow s): d_k=1} \alpha_{k-1}(s') \gamma_k^p(s', s) \beta_k(s)}{\sum_{(s' \rightarrow s): d_k=0} \alpha_{k-1}(s') \gamma_k^p(s', s) \beta_k(s)}$$

where

$$\gamma_k^p(s', s) \equiv P_r(R_{p_k} | p_k)$$

where  $p_k$  takes a value corresponding to  $(s', s)$ .

On receiving a signal sequence  $\mathbf{R}$ , the coder 1 calculates  $L_{ch}(d_k)$  and  $L_e(d_k)$  ( $k=1, 2, \dots, N$ ). The term  $L_a(d_k)$  ( $=L(d_k)$ ) necessary in the calculation is set to be zero. The decoder 1 sends  $L_e(d_k)$  to the decoder 2 via the interleaver. The decoder 2 calculates  $L_{ch}(d_n)$  and  $L_e(d_n)$  using the interleaved receive sequence  $\mathbf{R}'$  and  $L_a(d_n)$ , which is the  $L_e(d_n)$  received from the decoder 1. Completing the calculation, decoder 2 sends to decoder 1 the calculated  $L_a(d_n)$  via the deinterleaver as  $L_e(d_k)$ . The decoder 1 calculates again  $L_{ch}(d_k)$  and  $L_e(d_k)$  using the received signal sequence  $\mathbf{R}$  and  $L_a(d_k)$  which is  $L_e(d_k)$  sent from the decoder 2. As described earlier, only the newly updated  $L_e(d_k)$ , is sent to the other decoder to update the reliability information:  $L_{ch}(d_k)$  is common between the decoders and  $L_a(d_k)$  is received from the other decoder. This procedure of the calculation is repeated. After an appropriate number of calculations, decoder 2 makes decision as  $\hat{d}_n = 1(0)$  corresponding to  $\Lambda(d_n) > (<) 0$  and outputs data  $\hat{d}_k$  via the interleaver.

**The Max-Log-MAP Algorithm** To reduce computational complexity of the maximum a posteriori (MAP) decoding, the Max-Log-MAP and the soft output Viterbi algorithm (SOVA) are known [60]. Max-Log-MAP uses the following approximation:

$$\ln \left( \sum_i e^{x_i} \right) \approx \max_i (x_i)$$

To use the approximation, we define

$$A_k(s') \equiv \ln(\alpha_k(s')), \quad B_k(s) \equiv \ln(\beta_k(s)), \quad \Gamma_k(s', s) \equiv \ln(\gamma_k(s', s))$$

Applying to a recurrence Equation 7.98b for  $\alpha_k$ , we have

$$\begin{aligned} A_k(s) &= \ln \left( \sum_{\text{all } s'} \alpha_{k-1}(s') \gamma_k(s', s) \right) \\ &= \ln \left( \sum_{\text{all } s'} \exp[A_{k-1}(s') + \Gamma_k(s', s)] \right) \\ &\approx \max_{s'} \{A_{k-1}(s') + \Gamma_k(s', s)\} \end{aligned}$$



The maximum value is taken instead of the summation.  
Similarly, we get

$$\begin{aligned} B_{k-1}(s') &= \ln \left( \sum_{\text{alls}} \beta_k(s) \gamma_k(s', s) \right) \\ &= \ln \left( \sum_{\text{alls}} \exp[B_k(s) + \Gamma_k(s', s)] \right) \\ &\approx \max_s \{B_k(s) + \Gamma_k(s', s)\} \end{aligned}$$

Thus,  $A_k(s)$  and  $B_{k-1}(s')$  are calculated recursively, if  $\Gamma_k(s', s)$  is given.  $\Gamma_k(s', s)$  is given as

$$\begin{aligned} \Gamma_k(s', s) &\equiv \ln(\gamma_k(s', s)) \\ &= \ln[P_r(d_k) P_r(R_k | s_{k-1} s', s_k = s)] \\ &= \ln[P_r(d_k) P_r(R_{d_k} | d_k) P_r(R_{p_k} | p_k)] \end{aligned}$$

From Equation 7.101, we have

$$\Gamma_k(s', s) = c' + x_k L(d_k)/2 + \frac{E_b}{N_0 \rho} (R_{d_k} x_k + R_{p_k} y_k) \quad (7.102)$$

where  $c' = \ln(c_1 c_2 c_3)$  that never depends on  $x_k$  and  $y_k$  and therefore is omitted in the calculation. Consequently, the Max-Log-MAP algorithm reduces to the following approximate calculation:

$$\begin{aligned} L(d_k | \mathbf{R}) &= \ln \frac{\sum_{(s', s): d_k=1} \alpha_{k-1}(s') \gamma_k(s', s) \beta_k(s)}{\sum_{(s', s): d_k=0} \alpha_{k-1}(s') \gamma_k(s', s) \beta_k(s)} \\ &= \ln \frac{\sum_{(s', s): d_k=1} \exp[A_{k-1}(s') \Gamma_k(s', s) B_k(s)]}{\sum_{(s', s): d_k=0} \exp[A_{k-1}(s') \Gamma_k(s', s) B_k(s)]} \\ &\approx \max_{(s', s): d_k=1} \{A_{k-1}(s') \Gamma_k(s', s) B_k(s')\} - \max_{(s', s): d_k=0} \{A_{k-1}(s') \Gamma_k(s', s) B_k(s')\} \end{aligned} \quad (7.103)$$

The calculation is represented by a transition that gives the maximum value instead of summations over multiple transitions corresponding to  $d_k = a (a=0,1)$ . The logarithmic operations are avoided as in Equation 7.102, and the calculation complexity is reduced.

The SOVA, the other method for the calculation complexity reduction, introduces the two modifications to the conventional Viterbi algorithm. The first modification is the usage of the a priori probability  $P_r(d_k)$  in the branch metric calculations. The

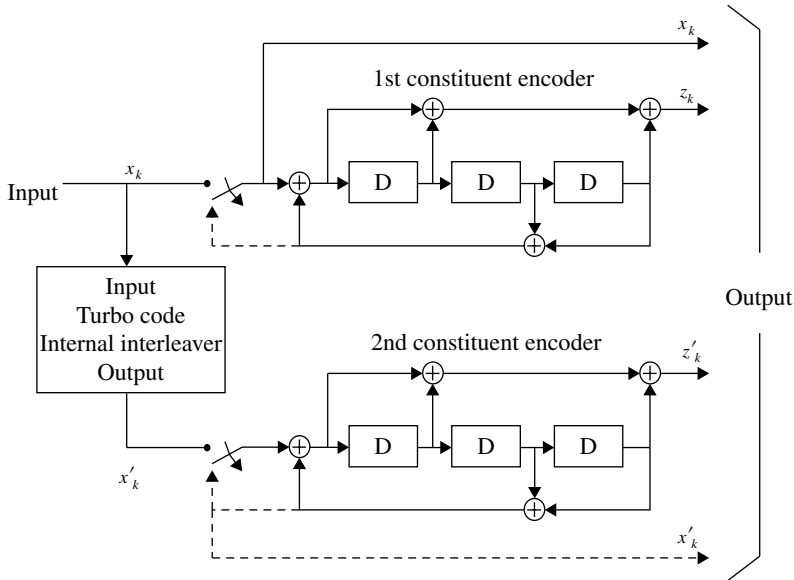


FIGURE 7.44 A standardized turbo code.

second modification is the output (soft) of the extrinsic reliability information to cope with the turbo iterative decoding. For further explanation, refer to [60].

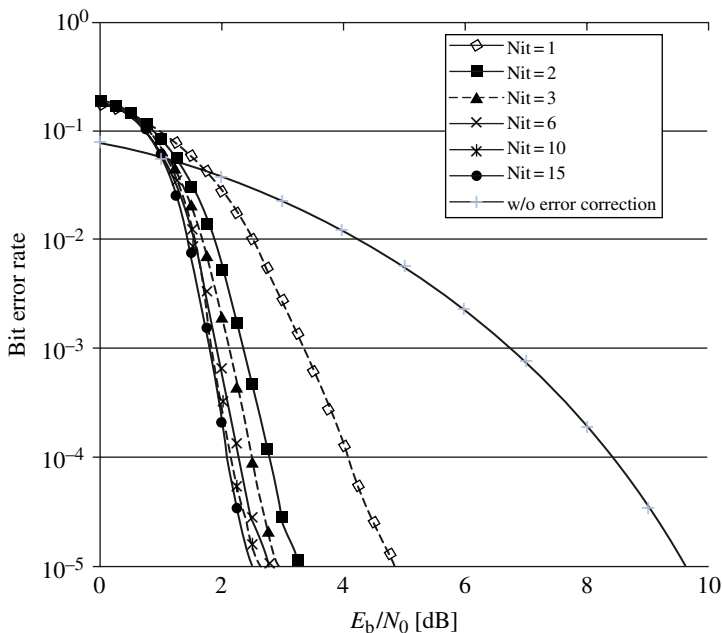
**An Example of Turbo Codes.** A turbo code that is standardized [62] in the third-generation cellular systems is shown in Figure 7.44 for an example. This is a systematic convolutional code with a code rate of  $1/3$ . Each coder consists of a feedback-type convolution circuit with three memory blocks.

The switch at the input of each coder is connected to a path shown with a solid line for a data period. After completion of the data input, the switches are connected to the path shown with a broken line, and signals  $x_k$ ,  $z_k$ ,  $z'_k$ , and  $x'_k$ , each of which takes three bits, are output.

By this operation, the signal state is terminated with  $S_N=0$ . Figure 7.45 shows bit error rate performance for an example. The decoding algorithm is the Max-Log-MAP, and a random interleaver is used. The bit error rate curves become steeper with increase of the iteration number. The curves show floors over the input  $E_b/N_0$  range. It is known that the values of the error rates floors depend on the size of the interleaver.

#### 7.4.6 LDPC Code

The LDPC (low density parity check) code was proposed by Gallager [63] in the 1960s. This is a linear block code, which has a parity check matrix whose elements scarcely take 1. The decoding is based on the MAP probability for each data bit. The code received little attention due to its high calculation complexity. The code is highlighted owing to its high error-correcting capability in the late 1990s [64].



**FIGURE 7.45** A turbo code performance example (code length  $n=960$ , constraint length  $k=4$ , and coding rate  $R=1/2$ ).

Here, we develop a brief description on the LDPC code. Consider a case, for example, when a parity check code is given as

$$\mathbf{H} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \quad (7.104)$$

From this, we see the code has a code length of  $N = 6$ , the number of the parity bits is  $K = 3$ , and therefore, the code rate is  $(N - K) / N = 0.5$ . The generator matrix  $\mathbf{G}$  for the code is given as a solution of the following equation:

$$\mathbf{G}\mathbf{H}^T = \mathbf{0} \quad (3 \times 3)$$

The solution is not uniquely obtained. Here, we take the solution for an example:

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Letting a transmit data  $\mathbf{d} = (d_1, d_2, d_3)$ , we have a transmit code  $\mathbf{c}$  as

$$\mathbf{c} = \mathbf{d}\mathbf{G}$$

To differentiate the transmit codes  $\mathbf{c}$ , we use a suffix as  $\mathbf{c}_i$  ( $i = 1, 2, \dots, 2^{N-k}$ ). We denote the received signals as  $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5, r_6)$ . If there is no error, the syndrome sequence is given by  $\mathbf{s} = \mathbf{r}\mathbf{H}^T = \mathbf{0}$  ( $1 \times 3$ ). When the parity check matrix  $\mathbf{H}$  is given with Equation 7.104, denoting  $\mathbf{s} = (s_1, s_2, s_3)$ , we have

$$\begin{aligned} s_1 &= r_1 \oplus r_2 \oplus r_3 \\ s_2 &= r_3 \oplus r_4 \\ s_3 &= r_4 \oplus r_5 \oplus r_6 \end{aligned}$$

where symbol  $\oplus$  means modulo 2 addition.

The MAP decoding gives the most probable value  $c_n = 0$  or  $c_n = 1$  of elements  $c_n$  of a code  $\mathbf{c} = (c_1, c_2, \dots, c_6)$  for a given received signal  $\mathbf{r}$ .

The a posteriori probability for  $c_n = a$  ( $a = 0$  or  $1$ ) is given with use of the Bayes' theorem as

$$P_r(c_n = a | \mathbf{r}) = \sum_{\mathbf{c}_i: c_{in}=a} P_r(\mathbf{r} | c_{in} = a) P_r(c_{in} = a) / P_r(\mathbf{r})$$

where  $\sum_{\mathbf{c}_i: c_{in}=a}$  means summation over all codes that takes  $c_n = a$ . We decide that  $\hat{c}_n = 1$  if  $P_r(c_n = 1 | \mathbf{r}) > P_r(c_n = 0 | \mathbf{r})$ , else  $\hat{c}_n = 0$ .  $P_r(\mathbf{r})$  is neglected in the calculation, since it never depends on  $c_n = a$ . Since each received signal elements are independent of each other, we have

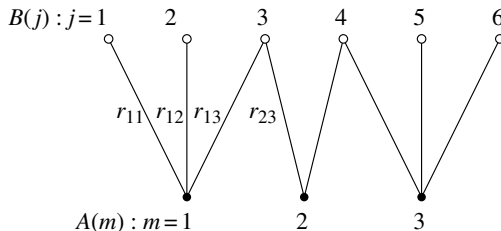
$$P_r(\mathbf{r} | c_{in} = a) P_r(c_{in} = a) = \prod_{n=1}^N P_r(r_n | c_{in} = a) P_r(c_{in} = a)$$

Therefore, we get

$$P_r(c_n = a | \mathbf{r}) = \sum_{\mathbf{c}_i: c_{in}=a} \prod_{n=1}^N P_r(r_n | c_{in} = a) P_r(c_{in} = a) / P_r(\mathbf{r})$$

Thus, the calculation becomes summation of products for probabilities. This expression is the same as that for the turbo decoding of a convolutional code, since both decoding belong to the MAP decoding. Therefore, the same method for an efficient calculation of the product/summation is adopted for both systems. The conditional probability  $P_r(\mathbf{r} | c_{in} = a)$  is determined by the received signal  $r_n$  if statistical properties of the noise are given. Under the Gaussian noise, it is given by equations similar to Equations 7.89 and 7.91.

A critical issue for the MAP decoding is to calculate the a priori probabilities  $P_r(c_{ij} = a)$ . In a feedback type convolutional coder, the transmit signal is affected by all the previous data. Therefore, all the multiplications in Equation 7.96 should be carried out. Contrary to this, for a block code, only the code elements that correspond to raw  $(1, 2, \dots, M)$  elements that take 1, of the error check matrix  $\mathbf{H}(M \times N)$ , are affected with each other in a way that the parity check results become zero. We denote the affected code element group by  $A(m)$ .



**FIGURE 7.46** The Tanner graph example.

For a case given with Equation 7.104, received signals that affect parity checks are schematically shown in Figure 7.46 (the Tanner graph [65]). In the figure,  $m$  corresponds to a row number and  $j$  corresponds to a column number of matrix  $\mathbf{H}$ , respectively. For  $m=1, j=1,2,3$ , with  $h_{1j}=1$ , are connected ( $A(1)=\{1,2,3\}$ ). Similarly, for  $m=2, j=3, 4$  ( $h_{2j}=1$ ) are connected ( $A(2)=\{3,4\}$ ), and for  $m=3, j=4,5,6$  ( $h_{3j}=1$ ) are connected ( $A(3)=\{4,5,6\}$ ).

When received signals are hardly decided, denoting the decided data by  $\hat{r}_j$  ( $j=1,2,\dots,6$ ), the syndrome is given with the parity check equations as

$$s_1 = \hat{r}_1 \oplus \hat{r}_2 \oplus \hat{r}_3$$

$$s_2 = \hat{r}_3 \oplus \hat{r}_4$$

$$s_3 = \hat{r}_4 \oplus \hat{r}_5 \oplus \hat{r}_6$$

Thus, in this figure,  $m$  expresses a group of the receive data  $r_j$  to which  $h_{mj}=1$  is multiplied in calculation of  $m$ th elements of a syndrome.

The LDPC decoding as well as the turbo decoding updates iteratively the a priori probabilities  $P_r(c_{ij} = a)$ .

To calculate the a priori probability, the external values are defined as follows:

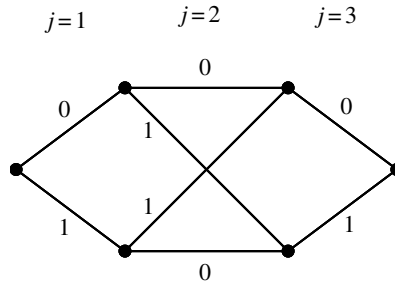
$$r_{mn}(0) = K \sum_{c_j} \prod_{c_n=0, j(\neq n) \in A(m)} q_{mj}(c_j) \Pr(r_j | c_j) \quad (7.105a)$$

$$r_{mn}(1) = K \sum_{c_j} \prod_{c_n=1, j(\neq n) \in A(m)} q_{mj}(c_j) \Pr(r_j | c_j) \quad (7.105b)$$

where  $\prod_{j(\neq n) \in A(m)}$  means taking products for running  $j$  included in  $A(m)$  except for  $j = n$ . Therefore,  $r_{mn}(a)$  ( $a=0,1$ ) never includes  $q_{mn}(a)$  and  $P_r(r_n | c_n)$ . The symbol  $\sum_{c_j: c_n=a}$  means summations over all combinations of the data connected to  $A(m)$ , which include  $c_n = a$ .

For examples of the data combination for the code given with Figure 7.44, regarding to  $A(1)(c_1 \oplus c_2 \oplus c_3 = 0)$ , we have combinations  $(c_2=0, c_3=0)$  and  $(c_2=1, c_3=1)$  for  $c_1=0$ , and  $(c_2=1, c_3=0)$  and  $(c_2=0, c_3=1)$  for  $c_1=1$ . We have similar combinations for  $c_2 = a$  and  $c_3 = a$ , ( $a=0,1$ ).

Constant value  $K$  is set to satisfy  $r_{mn}(0) + r_{mn}(1) = 1$ . The product/summations operations are carried out with the BCJR algorithm and the trellis presentation (Fig. 7.47) given under the parity check conditions as in decoding calculation of the turbo code.



**FIGURE 7.47** Trellis diagram for  $A(1)$  (even parity).

The iterative algorithm is summarized as follows:

Step 1: All of the a priori probabilities are initialized as  $q_{mn}(0) = q_{mn}(1) = 1/2$ .

Step 2:  $r_{mn}(0)$  and  $r_{mn}(1)$  are calculated with Equations 7.105a and 7.105b.

Step 3: Using  $r_{mn}(0)$  and  $r_{mn}(1)$  as external values, for every combinations of  $mn$  that correspond to  $h_{mn} = 1$ , we update  $q_{mn}(0)$  and  $q_{mn}(1)$  as follows:

$$q_{mn}(0) = K' \prod_{m'(\neq m) \in B(n)} r_{m'n}(0)$$

$$q_{mn}(1) = K' \prod_{m'(\neq m) \in B(n)} r_{m'n}(1)$$

where the multiplications are limited for  $m' \neq m$ . This fact means that the external values  $r_{mn}(a)$  from  $A(m)$  are never used in updating the a priori information  $q_{mn}(a)$ . The constants  $K'$  are determined so as to satisfy

$$q_{mn}(0) + q_{mn}(1) = 1$$

Step 4 : (Code word estimation) For  $n=1, 2, \dots, N$ , we calculate

$$Q_n(0) = P(r_n | c_n = 0) \prod_{m \in B(n)} r_{mn}(0)$$

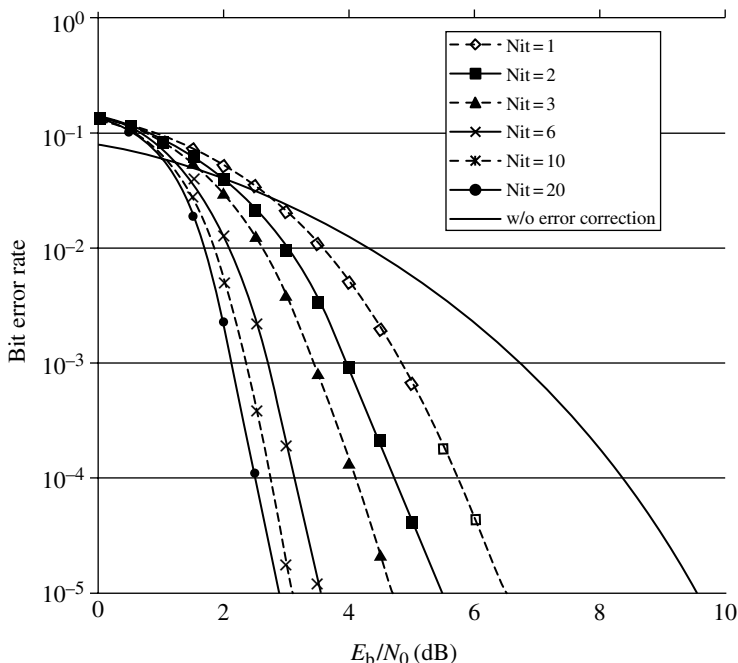
$$Q_n(1) = P(r_n | c_n = 1) \prod_{m \in B(n)} r_{mn}(1)$$

This means the calculation of joint probabilities of  $r_n$  and  $c_n = a$ , that is,  $P(r_n, c_n = a) = P(r_n | c_n = a)P(c_n = a)$ , using, for  $c_n$ , the a priori probabilities  $P(c_n = a)$ , which is given by the products of all the external values as  $P(c_n = a) = \prod_{m \in B(n)} r_{mn}(a)$ . We estimate as  $\hat{c}_n = 0$  if  $Q_n(0) > Q_n(1)$ ; else we estimate as  $\hat{c}_n = 1$ .

Step 5: (Parity check) If syndromes are calculated as

$$\mathbf{s} = (\hat{c}_1, \hat{c}_2, \dots, \hat{c}_N) \mathbf{H}^T = \mathbf{0}$$

then the estimated word is output and the algorithm terminates, otherwise, we go back to step 2. If the syndromes never become zero even after a fixed number of iterations, the word  $\hat{\mathbf{c}}$  estimated so far is the output and the calculation procedure stops.



**FIGURE 7.48** LDPC code performance example.

This algorithm is called the sum-product method. The other method is based on the calculation of the log-likelihood as discussed in the turbo decoding system.

It is shown that the LDPC codes with a long code length and proper setting of the parity check matrix, and therefore the code words, yield bit error rate performance close to the Shannon limit. The property of the sparseness of the parity check matrix (the matrix elements take scarcely “1”) is the key point for the good error rate performance while suppressing the increase of computational complexity. Examples of bit error rate performance of an LDPC code are shown in Figure 7.48.

#### 7.4.7 A Phenomenological Expression of the A Priori Probability and Error Rates

The a priori probability  $\lambda \left( = \ln \frac{P_r(d_k=1)}{P_r(d_k=0)} \right)$  is given to a receiver decoder from another decoder as extrinsic probability. Estimation accuracy of the a priori probability is evolved through iterative exchange of the extrinsic probability between the two decoders. The estimated a priori probability depends on the codes and its decoding methods. It is difficult to give error rates in an iterative decoding process for a given code, although some theoretical analyses of the iterative estimation are presented [66–68]. Here, we discuss a phenomenological expression of the error rates for a code by using several macroscopic parameters, which are extracted from experimental results of error rates of the codes.

Denoting probability density function of  $\lambda$  by  $f(\lambda)$ , we have average error rates as (Section 3.3.2)

$$\langle P_e \rangle = \int_{-\infty}^{\infty} Q \left[ \sqrt{2\gamma} \left( 1 + \frac{\lambda}{4\gamma} \right) \right] f(\lambda) d\lambda \quad (7.106)$$

For a system with the output symmetry and assumption of all of the data of 1, the following symmetry relation is known [68]:

$$\ln \frac{f(\lambda)}{f(-\lambda)} = \lambda \quad (7.107)$$

Following references [66, 67], we approximate  $f(\lambda)$  with the Gaussian distribution as

$$f(\lambda) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(\lambda-\mu_n)^2}{2\sigma_n^2}} \quad (7.108)$$

where we introduced suffix  $n$  to show  $n$ th iteration. Following usual notation of the signal to noise power ratio (SNR), we define the following SNR:

$$\gamma_n = \frac{\mu_n^2}{\sigma_n^2}$$

From Equation 7.107, we have  $2\mu_n = \sigma_n^2$  and therefore

$$\mu_n = 2\gamma_n \sigma_n = 2\sqrt{\gamma_n}$$

Thus, Equation 7.108 is expressed with  $\gamma_n$  as

$$f(\lambda) = \frac{1}{\sqrt{8\pi\gamma_n}} e^{-\frac{(\lambda-2\sqrt{\gamma_n})^2}{8\gamma_n}}$$

We try to express  $\gamma_n$  with  $E_b/N_0 (= \gamma)$ . Conditions required qualitatively for  $\gamma_n$  are as follows: (i) The probability density function  $f(\lambda)$  should shift toward a higher  $\lambda$  region with increase of  $\gamma$ , since  $\lambda$  is estimated LLR and the estimation accuracy increases with  $\gamma$ . (ii)  $\gamma_n$  should increase with  $n$ . However, this is valid for  $\gamma$  higher than the Shannon limit threshold  $\gamma_c$ . To satisfy these conditions, we adopt an empirical equation as

$$\gamma_n = a_n \left( \frac{\gamma - \gamma_c}{\gamma_{n0} - \gamma_c} \right)^{m(n)} \quad \gamma > \gamma_c$$

Parameters  $a_n$ ,  $m(n)$ , and  $\gamma_{n0}$  are determined by curve fitting with experimental error rates.



TABLE 7.2    Extracted Parameters

$n$	$m(n)$	$\gamma_{n0}(\text{dB})$	$a_n$
1	1	4	4.5
2	1.7	2.2	7.3
3	2	1.5	8.8
6	3	0.95	10
18	6	0.7	10.8

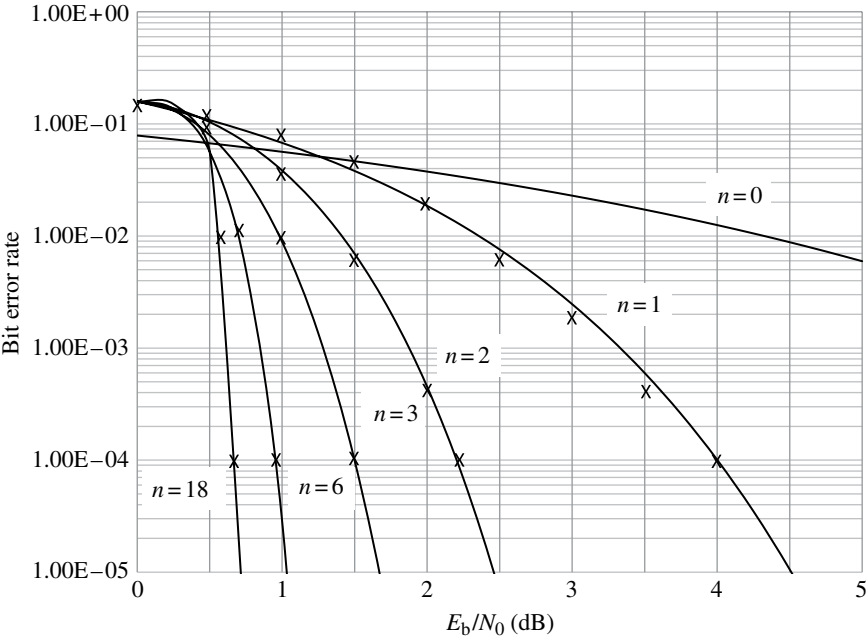


FIGURE 7.49    Calculated error rate performance for a turbo code (marked with  $x$  shows experimental results).

**A Calculation Example** Table 7.2 shows extracted parameters given by curve fitting with results given by Berrou et al. [58]. Their system uses a code with a rate of 1/2 and two-level binary baseband signal. From these facts and the Shannon channel capacity formula, we have

$$\log_2 \left( 1 + \frac{S}{N} \right) = \log_2 \left( 1 + \frac{E_b}{N_0} \right) = \frac{C}{W} = 1$$

From this result, we have  $\gamma_c = 0$  [dB].  $\gamma_{n0}$  is determined as a  $\gamma$  value that gives an error rate of  $10^{-4}$ . The calculated bit error rates using these parameters and Equation 7.106 are shown in Figure 7.49. The points marked with cross in the figure show the experimental values chosen representatively. The calculated results agree fairly with the experimental results in the considered region of  $E_b/N_0$ .

### 7.4.8 ARQ

A system employing an ARQ scheme requests retransmission of the data received in error. There are three ARQ strategies: (1) Stop and Wait, (2) Go Back  $N$ , and (3) Selective Repeat. In the Stop and Wait ARQ system, the receiver sends an *ACK* signal to acknowledge a correct reception and sends a *NAK* signal to request a retransmission of a coded signal block detected in error. The efficiency of transmission is low, since confirmation of the received signal is made for every block of data.

In the Go Back  $N$  ARQ system, coded data block are transmitted continuously. If the sender receives a *NAK* signal requesting a retransmission, then the transmitter goes back by  $N$  data blocks and restarts the transmission from that data block. The time delay between the transmission of a data block and the reception of a *NAK* signal must be less than the time for the transmission of  $N$  data blocks. The idle time is shorter for this system than for the Stop and Wait ARQ system. However, it still wastes time since  $N$  data blocks, some of which may be without errors, are retransmitted for every detected error.

In the Selective Repeat ARQ system, data blocks are sent continuously and only the blocks in error are retransmitted. Thus, this system has the highest efficiency. The management of the order of data blocks and the buffer is more complicated for this system compared to the other systems.

Although both FEC and ARQ cope with receive errors, their effects on the errors differ from each other. The FEC tends to fail to correct errors and to output erroneous bits when error rate increases over its error correcting capability or burst errors occur. Code rate should be lowered to enhance the error correcting capability. Therefore, channel usage efficiency drops even in a channel with good quality. Contrary to the FEC, ARQ is ineffective for random errors resulting into increase of retransmission. However, it is efficient against burst errors.

Probabilities of error detection failure for ARQ are neglected in actual situation. Furthermore, decrease of channel usage efficiency due to parity check bits is kept small for a long data block, since the parity check bit length is fixed such as 16 or 32 bits in CRC codes (Section 7.4.2). As discussed earlier, FEC and ARQ are complementary to each other as a countermeasure against transmission errors. Therefore, a joint use of FEC and ARQ is effective, and it is known as H(hybrid)-ARQ. In this system, parity bits for error detection are always sent together with data bits. Hereafter, we include the parity bits into data bits for simplicity.

The H-ARQ is grouped into type I and type II as follows. The type I always sends FEC code word and continues to send it until request for retransmission stops. The type II sends only data bits for the first time. If error is not detected, the next data are sent. If error is detected and request for retransmission is received, the FEC parity check bits for those data are sent. If receiver fails to correct the errors even after receiving the FEC parity check bits, further strategies are different on H-ARQ systems. The type II method, which sends only data bits at the first transmission, is advantageous under a good channel condition that causes errors scarcely. The type I method is advantageous in view of less retransmission times, even for a high error rate channel as long as errors are successfully corrected with FEC. In this channel condition, a decrease of channel efficiency for type II is kept small since only the parity bits are sent on a request.

Retransmission stops after no error is detected at the receiver. To reduce retransmission times, the methods are known as soft combining ARQ, where received data packet is stored and is combined with newly received data packet. A method is known as the Chase combining, where the same packet is sent multiple times and MRC (Section 7.1.1) of each symbol is carried out. The symbol to noise power ratio  $E_s/N_0$  increases with the maximal ratio combining. The other method is called incremental redundancy (IR), where different codes are sent at each retransmission: for example, puncturing position of parity bits in a systematic convolutional code is changed for each retransmission. Error-correcting capability improves owing to increased number of useful parity bits. For the systematic code, data bits are combined with the maximum ratio algorithm. The coding gain increases with a penalty of low coding rate (or channel efficiency). The effect of the coding gain is higher than that of increased  $E_s/N_0$  owing to the MRC, and therefore, IR systems show better error rate performance than the Chase combining.

For further discussion, refer to [69]. H-ARQ is used for systems such as HSPA and LTE (Section 9.7.6) [70].

#### 7.4.9 Applications to Mobile Radio Channels

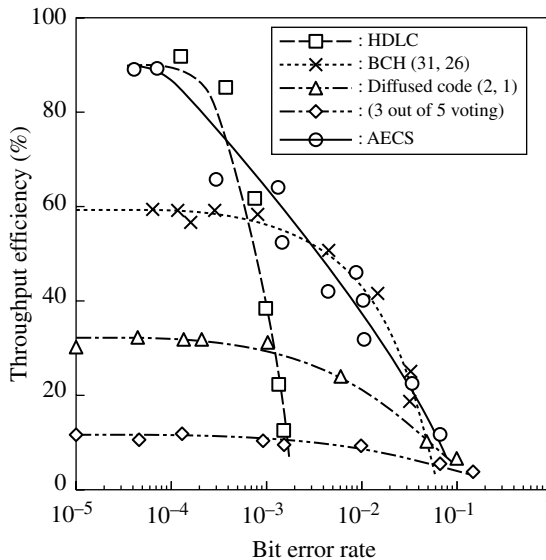
Mobile radio channels are characterized by burst errors and high average error rates caused by fading. Error control techniques can be used depending on the requirements of the application. Error detection is a must for data services.

Burst errors correcting codes or interleaving techniques are used to cope with burst error. These techniques are useless for a very long burst error. In this case, an ARQ system works effectively.

A data transmission system, which automatically changes its mode of transmission (adaptive error control system (AECS)), is developed in [71] for application to analog cellular systems. A transmission rate of 4800 kbps is possible with a voice band data modem (modulator/demodulator) (CCITT standard V.27 ter.). The error rate performance is shown in Figure 7.50. For error rates less than  $10^{-4}$ , Selective Repeat ARQ without error correction (HDLC) is used in view of transmission efficiency. For error rates around  $10^{-3}$ , a 3-error correcting BCH code ( $n=31, k=26$ ) is used with interleaving. For errors of rate of around  $10^{-2}$ , a burst error correcting code with (diffused code) a code rate of  $1/2$  is used. For high error rates such as  $10^{-1}$ , five successive transmissions of a data block are made. At the receiver, majority voting decoding is performed. Thus, the code rate with the above modes of error correction varies as  $26/31=0.84$ ,  $0.5$ , and  $0.2$ .

### 7.5 TRELLIS-CODED MODULATION

Trellis-coded modulation [72, 73] is a combined coding and modulation technique: coding and modulation are integrated to achieve better error rate performance without compromising bandwidth. To understand the principle of this technique, it is helpful to consider a system where coding and modulation are introduced independently. Let



**FIGURE 7.50** Simulated throughput efficiency versus bit error rate for coded data transmission in analog cellular channels. Courtesy of OKI Electric Co. Ltd.

us compare a QPSK system without coding and an 8PSK system with a coding rate of 2/3. For the same data rate, both systems have the same spectral bandwidth. In the 8PSK receiver, demodulation and symbol-by-symbol decision-making are performed, followed by the error correction process. The 8PSK system yields no significant advantage in error rate performance over the QPSK system without coding. This is because the coding gain is lost due to the higher error probability in the 8PSK system.

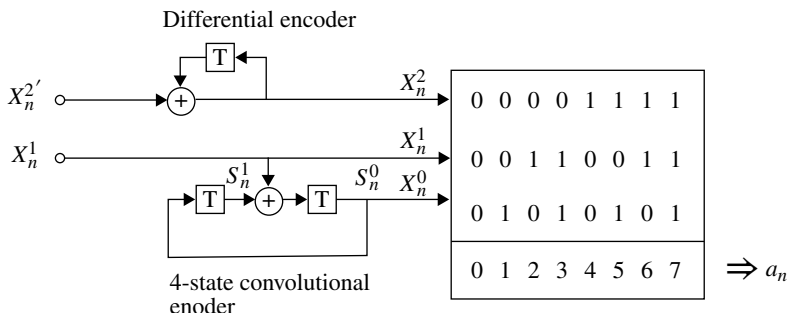
In trellis-coded modulation, coding and modulation are combined to produce the minimum error rates provided that the signals are detected as a sequence: the MLSE. The criterion is to minimize Euclidean distance between coded and modulated signal sequences. For signal sequences with the minimum distance, called the free distance  $d_{\text{free}}$ , the uncorrelated Gaussian noise with variance  $\sigma^2$ , error rates are approximately given as

$$P_e \approx N_{\text{free}} Q(d_{\text{free}}/2\sigma) \quad (P_e \ll 1) \quad (7.109)$$

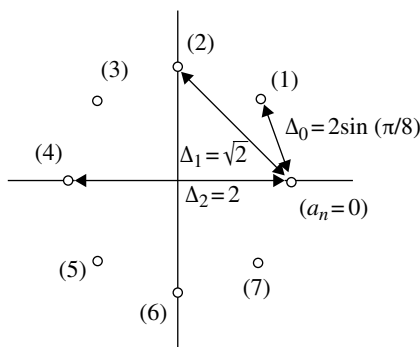
where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{y^2}{2}\right) dy$$

and  $N_{\text{free}}$  denotes the average number of nearest-neighbor signal sequences with distance  $d_{\text{free}}$  that diverge at any state from a transmitted signal sequence and remerge with it after one or more transitions.



**FIGURE 7.51** Trellis diagram for 4-state trellis-coded 8PSK [74]. Reproduced by permission of © 1987, IEEE.

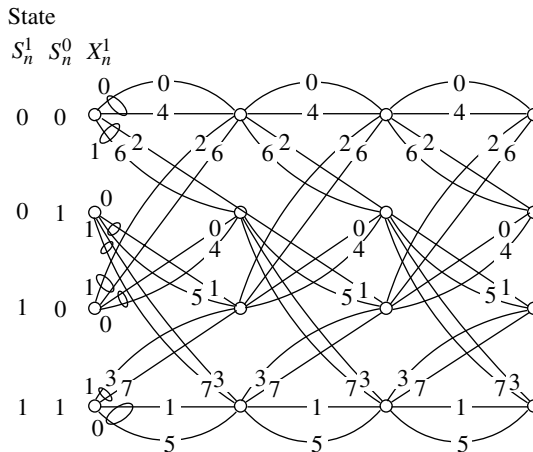


**FIGURE 7.52** 8-PSK signal.

As an example of a trellis-coded modulation system, an encoder for an 8PSK code is shown in Figure 7.51. A bit of the di-bit input signal is fed to a 4-state convolutional encoder with a coding rate of 1/2. Thus, the total coding rate is 2/3. The code word  $a_n$  is assigned a phase as in Figure 7.52. The 4-state trellis diagram for the trellis-coded 8PSK system is shown in Figure 7.53. We consider (Euclid) distance normalized with the radius  $A$  of the signal constellation. The 8PSK signals are assigned to the transitions in the 4-state trellis according to the following rules:

- I. Parallel transitions occur with signals in the subsets (0,4), (1,5), (2,6), or (3,7). The distance between signals in this subset is at least  $\Delta_2 = 2$ .
- II. Four transitions originating from or merging in one state are from signals in the subsets (0,4,2,6) or (1,5,3,7). The distance between the signal in this subset is  $\Delta_1 = \sqrt{2}$ .

Any two signal sequences in the trellis of Figure 7.53 originating from one state and remerging in another state after more than one transition have a squared distance of at least  $\Delta_1^2 + \Delta_0^2 + \Delta_1^2 = \Delta_0^2 + \Delta_2^2$  between them. Since the squared distance between the parallel paths is  $\Delta_2^2$ , the free distance  $d_{\text{free}}$  in the 4-state 8PSK system



**FIGURE 7.53** Trellis diagram for 4-state trellis-coded 8PSK [74]. Reproduced by permission of © 1987, IEEE.

becomes  $d_{\text{free}} = 2A$ . If we compare this with the  $d_{\text{free}} = \sqrt{2} A$  for uncoded QPSK, a 3 dB gain is obtained for the trellis-coded 8PSK over the uncoded QPSK. More gains are achieved with other codes. Trellis-coded modulation is applied to other modulation schemes such as QAM.

Trellis coding assumes coherent detection and MLSE, for example, the Viterbi algorithm.

The applications of trellis-coded modulation to fading channels are discussed in [75] and [76]. A trellis-coded 16QAM system has been investigated for applications in TDMA systems for mobile radio communications [77]. A trellis-coded 8PSK system with envelope control (for power efficiency in the mobile radio transmitter amplifier) is described in [78].

## 7.6 ADAPTIVE INTERFERENCE CANCELLATION

In radio communication systems, the received signal is sometimes interfered with by other signals. The interference may be intentional or unintentional. Cancellation of the intentional interference (“jamming”) in communication or radar systems is important in military applications. Even in nonmilitary systems, interference cancellation becomes important since it results in a more efficient use of the spectrum, which is important in cellular systems. Interference cancellation techniques for mobile radio communications have been a hot topic.

Interference cancellation must be adaptive since both the interfering signal and the desired signal are time varying in mobile radio communications. There are two categories of adaptive interference cancellation techniques. One uses adaptive antenna arrays [79] and the other uses signal processing techniques to suppress the interfering signal.

### 7.6.1 Adaptive Array Antenna

Let us consider adaptive antennas first. Adaptive antennas are used to induce a null point in the directive pattern in the direction of the interfering signal. Consider, for example, an antenna array system consisting of two antennas that has an omnidirectional directive pattern in horizontal plane as shown in Figure 7.54. The received signals are weighted with complex coefficients and summed.

Let us assume a plane wave for simplicity. The signals at antennas 1 and 2 become the same except for a phase difference  $\Delta\varphi$  that is  $\Delta\varphi = (2\pi d/\lambda) \cos\theta$ ;  $d$  is the distance between the antennas and  $\lambda$  is the wavelength of the carrier. The effect the modulating signal has on the received signal is neglected: this is valid when the modulating signal changes only little in the arrival time difference between the signals. Let the complex weighting factors  $w_1$  and  $w_2$  satisfy  $|w_1| = |w_2| = 1$  and  $\angle w_1 - \angle w_2 = \Delta\theta$ ; then the directive pattern of the antenna array or the relative level of the summed signal as a function of  $\theta$  becomes

$$G(\theta) = \cos^2 \left( \frac{\pi d}{\lambda} \cos\theta + \frac{\Delta\theta}{2} \right) \quad (7.110)$$

A null point is produced in the direction of  $\theta_0$  by letting  $\Delta\theta = -(\pi d/\lambda) \cdot \cos\theta_0 \pm \pi/2$ . A directive pattern is shown in Figure 7.55 for  $d = \lambda/2$  and  $\Delta\theta = (1 - 1/\sqrt{2})\pi$ . By using  $n$  antennas, we can have  $2(n-1)$  null points, of which  $n-1$  are independent.

In order for the antenna array to be adaptive to the interference, the weighting factors  $w_1$  and  $w_2$  are automatically adjusted. A well-known method is the LMS

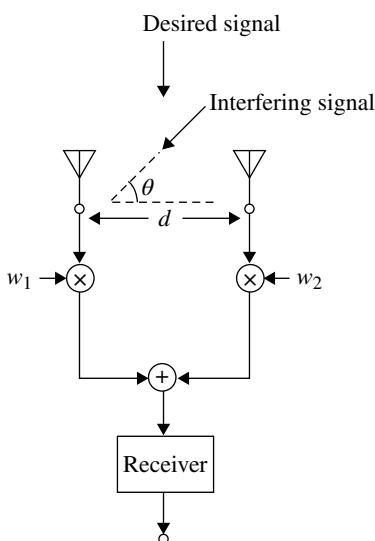
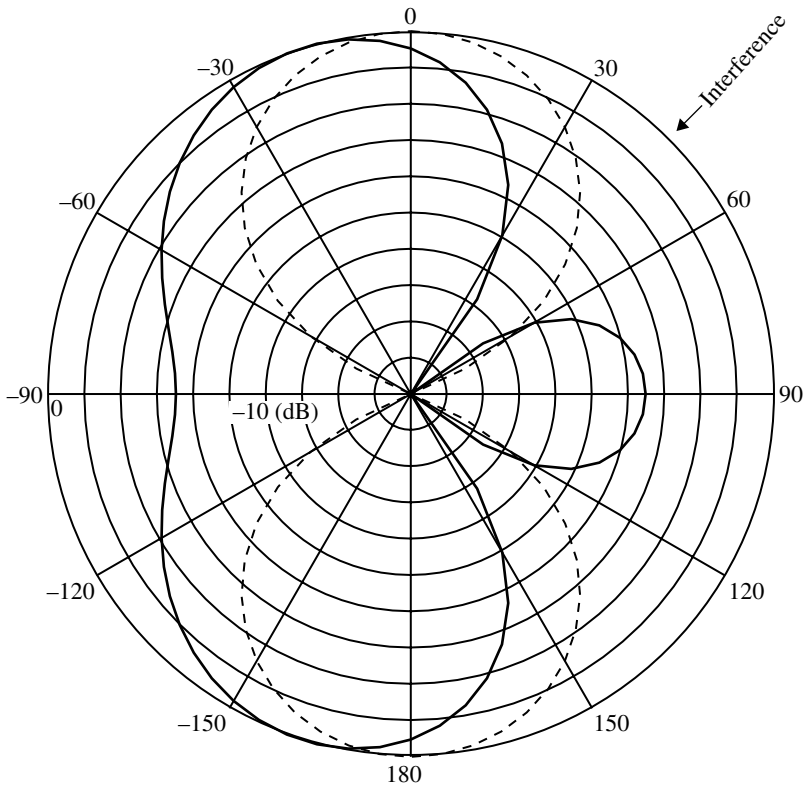


FIGURE 7.54 An antenna array.



**FIGURE 7.55** Directive pattern of an antenna array.

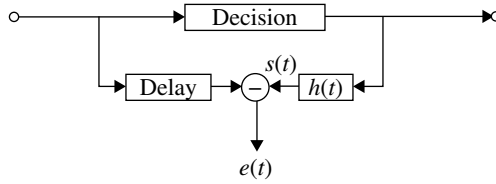
algorithm (Section 2.5.2). The weighting factors are adjusted to minimize the average LMS error

$$\langle |e(t)|^2 \rangle = \left\langle \left| s(t) - \sum_{i=1}^N w_i r_i(t) \right|^2 \right\rangle$$

$r_i(t)$  is the received signal at antenna  $i$ , and  $s(t)$  is the reference signal or the desired signal.

In radar systems, the reference signal is given; it is a transmitted and reflected signal. In analog communication systems, the reference signal must be transmitted periodically. In digital communication systems, the reference signal can be generated from the received signal as shown in Figure 7.56;  $h(t)$  is the impulse response of the overall channel including transmit and receive filters. The capability of digital systems to generate the reference signal is owing to discrete-level signal transmission. To get the optimum weighting factors, recursive methods can be applied as discussed in Section 7.3 for the automatic equalizer. A diversity receiver combined with the LMS algorithm acts as an adaptive interference canceller if the interference levels become significant and as an MRC diversity system if the interference is absent [80, 81].





**FIGURE 7.56** Generation of an error signal for digital communication system.

**Optimum Weight and Attainable Signal to Interference plus Noise Power Ratio** Now, we will obtain the optimum weights of an adaptive antenna and the maximum signal to interference plus noise power ratio (SINR) based on the MMSE criterion under the existence of interference and noise. We assume that  $M$  signals  $s_i$  ( $i=1, 2, \dots, M$ ) including a desired signal are received at  $N$  antennas and noise  $n_n$  ( $n=1, 2, \dots, N$ ) are added. The signal and noise at  $n$ th antenna are given as

$$r_n = \sum_{j=1}^M h_{nj} s_j + n_n \quad (n=1, 2, \dots, N)$$

where  $h_{nj}$  denotes propagation coefficient for transmit signal  $s_j$  from transmitter to receive antenna  $n$ . The desired signal is denoted by  $s_i$ . The output signal of the adaptive antenna with weighting factor  $w_n$  is given as

$$\hat{s}_i = \sum_{n=1}^N w_n r_n$$

The mean square error is expressed as

$$J = \langle |s_i - \hat{s}_i|^2 \rangle$$

where  $\langle \cdot \rangle$  symbol means ensemble average. We assume for signals and noises as

$$\begin{aligned} \langle s_i s_j^* \rangle &= \begin{cases} S & (i=j) \\ 0 & (i \neq j) \end{cases} \\ \langle n_i n_j^* \rangle &= \begin{cases} N & (i=j) \\ 0 & (i \neq j) \end{cases} \\ \langle s_i n_j^* \rangle &= s_i^* n_j = 0 \end{aligned} \quad (7.111)$$

To obtain optimum weights  $w_m$  that give the minimum average square error  $J$ , we differentiate  $J$  with respect to  $w_m^*$  as

$$\frac{\partial}{\partial w_m^*} J = 0 \quad (m=1, 2, \dots, N) \quad (7.112)$$

We get

$$\frac{\partial}{\partial w_m^*} |s_i - \hat{s}_i|^2 = -(s_i - \hat{s}_i) r_m^*$$

From the above result and Equation 7.111, we have

$$\langle s_i r_m^* \rangle = h_{mi}^* \langle |s_i|^2 \rangle = h_{mi}^* S \quad (7.113)$$

$$\begin{aligned} \langle \hat{s}_i r_m^* \rangle &= \left\langle \left( \sum_{n=1}^N w_n r_n \right) r_m^* \right\rangle \\ &= \sum_{n=1}^N w_n \left[ \sum_{j=1}^M h_{nj} h_{mj}^* S + N \delta_{mn} \right] \left( \delta_{mn} = \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases} \right) \end{aligned} \quad (7.114)$$

Using Equations 7.112–7.114, we get a matrix expression as

$$[\mathbf{H} + \rho^{-1} \mathbf{I}] \mathbf{w}^T = \mathbf{h}_i^{*T} \quad (\mathbf{T} : \text{transpose})$$

where  $\mathbf{w} = (w_1, w_2, \dots, w_N)$ ,  $\mathbf{h}_i = (h_{1i}, h_{2i}, \dots, h_{Ni})$ , elements of matrix  $\mathbf{H}$  are  $H_{mn} = \sum_{j=1}^M h_{nj}^* h_{mj}$ ,  $\rho = S / N$ , and  $\mathbf{I}$  ( $N \times N$ ) is a unit matrix. Solving the above equation, we get the optimum weights as

$$\mathbf{w}^T = [\mathbf{H} + \rho^{-1} \mathbf{I}]^{-1} \mathbf{h}_i^{*T}$$

Then the desired signal power  $S_o$ , interference power  $I_o$ , and noise power  $N_o$  are given as

$$\begin{aligned} S_o &= \left\langle \left| \sum_{n=1}^N w_n h_{nj} s_i \right|^2 \right\rangle = |\mathbf{h}_i \mathbf{w}^T|^2 S \\ I_o &= \left\langle \left| \sum_{n=1}^N w_n h_{nj} s_i \right|^2 \right\rangle = \sum_{j=1, j \neq i}^N |\mathbf{h}_j \mathbf{w}^T|^2 S \\ N_o &= \mathbf{w} \mathbf{w}^{*T} N = \|\mathbf{w}\|^2 N \end{aligned} ,$$

The maximum SINR becomes

$$\text{SINR} = \frac{S_o}{I_o + N_o} = \frac{|\mathbf{h}_i \mathbf{w}^T|^2}{\sum_{j=1, j \neq i}^N |\mathbf{h}_j \mathbf{w}^T|^2 + \|\mathbf{w}\|^2 \rho^{-1}} \quad (7.115)$$

So far, we assumed that all of the propagation coefficients are known. If they are unknown, the weights are given by solving the following equation:

$$\sum_{n=1}^N \langle r_n r_m^* \rangle w_n = \langle s_i r_m^* \rangle \quad (m = 1, 2, \dots, N)$$

When we get correlations,  $\langle r_n^* r_m^* \rangle$  and  $\langle s_i^* r_m^* \rangle$ , the accuracy of estimation decreases and consequently the attainable SINR drops, if the number of samples is not sufficient. The propagation coefficients  $h_{ni}$  for the desired signal are usually estimated with reference signals to demodulate the signal. In this case, the estimation accuracy of the correlations can be enhanced by using  $h_{ni}$  as follows.

We subtract the desired signal from received signals. At antenna  $n$ , we let

$$r_n' = r_n - h_{ni} s_i$$

Then we have

$$\langle r_n' r_m'^* \rangle = \sum_{j=1, j \neq i}^M \langle s_{nj} s_{mj}^* \rangle + N \delta_{nm}$$

where  $s_{nj} = h_{nj} s_j$ . Estimation accuracy of  $\langle r_n' r_m'^* \rangle$  is enhanced, since the desired signal is removed. Using  $\langle r_n' r_m'^* \rangle$ , we get

$$\langle r_n^* r_m^* \rangle = h_{ni}^* h_{mi} S + \langle r_n' r_m'^* \rangle$$

Using Equation 7.113, we have

$$\langle s_i^* r_m^* \rangle = h_{mi}^* S$$

**Discussion** In the following, we show that the receiver with the MMSE criterion becomes interference canceller and MRC diversity system for the absence of noise or interference, respectively. For simplicity, we assume a system with two antenna elements. Then the matrix  $\mathbf{H} + \rho^{-1} \mathbf{I}$  becomes

$$\mathbf{H} + \rho^{-1} \mathbf{I} = \begin{pmatrix} |h_{11}|^2 + |h_{12}|^2 + \rho^{-1} & h_{11}^* h_{21} + h_{12}^* h_{22} \\ h_{11} h_{21}^* + h_{12} h_{22}^* & |h_{21}|^2 + |h_{22}|^2 + \rho^{-1} \end{pmatrix}$$

(i) *In Case of Absence of Noise.* Putting  $\rho^{-1} = 0$  in the above equation, we have the matrix

$$\begin{aligned} \mathbf{H} &= \begin{pmatrix} |h_{11}|^2 + |h_{12}|^2 & h_{11}^* h_{21} + h_{12}^* h_{22} \\ h_{11} h_{21}^* + h_{12} h_{22}^* & |h_{21}|^2 + |h_{22}|^2 \end{pmatrix} \\ &= \begin{pmatrix} h_{11}^* & h_{12}^* \\ h_{21}^* & h_{22}^* \end{pmatrix} \begin{pmatrix} h_{11} & h_{21} \\ h_{12} & h_{22} \end{pmatrix} \end{aligned}$$

From  $\mathbf{H} \mathbf{w}^T = \mathbf{h}_i^{*T}$  and multiplying  $\begin{pmatrix} h_{11}^* & h_{12}^* \\ h_{21}^* & h_{22}^* \end{pmatrix}^{-1}$  from the left-hand side, we get

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \equiv \begin{pmatrix} h_{11} & h_{21} \\ h_{12} & h_{22} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} h_{11}^* & h_{12}^* \\ h_{21}^* & h_{22}^* \end{pmatrix}^{-1} \begin{pmatrix} h_{11}^* \\ h_{21}^* \end{pmatrix}$$

Since  $\begin{pmatrix} h_{11}^* & h_{12}^* \\ h_{21}^* & h_{22}^* \end{pmatrix}^{-1} = \frac{1}{\Delta} \begin{pmatrix} h_{22}^* & -h_{12}^* \\ -h_{21}^* & h_{11}^* \end{pmatrix}$  where  $\Delta = h_{11}^* h_{22}^* - h_{12}^* h_{21}^*$ , we obtain

$$y_2 = w_1 h_{12} + w_2 h_{22} = \frac{1}{\Delta} (-h_{21}^* h_{11}^* + h_{11}^* h_{21}^*) = 0$$

The interference signal component is given as  $y_2 s_2$ , and therefore, the interference is shown to be cancelled.

(i) *In Case of Absence of Interference.* Putting  $h_{12} = h_{22} = 0$ , we have

$$\mathbf{G} \equiv \mathbf{H} + \rho^{-1} \mathbf{I} = \begin{pmatrix} |h_{11}|^2 + \rho^{-1} & h_{11}^* h_{21} \\ h_{11} h_{21}^* & |h_{21}|^2 + \rho^{-1} \end{pmatrix}$$

Then the optimum weights are given as

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \mathbf{G}^{-1} \begin{pmatrix} h_{11}^* \\ h_{21}^* \end{pmatrix}$$

Using

$$\mathbf{G}^{-1} = \frac{1}{\Delta} \begin{pmatrix} |h_{21}|^2 + \rho^{-1} & -h_{11}^* h_{21} \\ -h_{11} h_{21}^* & |h_{11}|^2 + \rho^{-1} \end{pmatrix} \text{ where } \Delta = \rho^{-1} (|h_{11}|^2 + |h_{21}|^2) + \rho^{-2}$$

we get

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \frac{\rho^{-1}}{\Delta} \begin{pmatrix} h_{11}^* \\ h_{21}^* \end{pmatrix}$$

From the arguments in Section 7.2.1 (Eq. 7.49a), the above result shows the optimum weights are obtained as the MRC diversity receiver.

**Performance Example.** Here, we analyze receiver performance with interference cancellation under simplified conditions. The system model is shown in Figure 7.54. An array antenna with omnidirectional two elements is assumed under plane wave propagation condition. The desired signal  $s$  arrives perpendicular to the line of the two antenna elements. The incident angle  $\theta$  of the interfering signal  $i$  is varied. Combining diversity system, interference canceller and minimum mean square receiver are considered. The propagation constants at the two antenna elements differs by a factor of  $e^{-j2\pi d \cos \theta / \lambda}$ .

The receiver output signal  $r$  becomes

$$r = w_1 (s + i + n_1) + w_2 (s + e^{-j\Delta\theta} i + n_2)$$

where  $\Delta\theta = 2\pi d \cos \theta / \lambda$ ,  $n_1$  and  $n_2$  are added noises, and the weighting factors  $w_1$  and  $w_2$  are assumed to be normalized as  $|w_1|^2 + |w_2|^2 = 2$  for simplicity.

*Combining Diversity:* The weights become  $w_1 = w_2 = 1$ . The output SINR becomes

$$\text{SINR} = \frac{S}{I \cos^2(\Delta\theta / 2) + N/2} \quad (\text{diversity})$$

where  $S$ ,  $I$ , and  $N$  are the input signal, interference, and noise power per antenna element, respectively.

*Interference Canceller:* The weights are given as  $w_2 = w_1 e^{-j\Delta\theta}$ . The output SNR becomes

$$\text{SNR} = \frac{2S}{N} \sin^2(\Delta\theta / 2) \quad (\text{canceller})$$

*The MMSE Receiver:* We define the average mean square error as

$$J = \langle |2s - r|^2 \rangle$$

Making  $\frac{\partial}{\partial w_1^*} J = \frac{\partial}{\partial w_2^*} J = 0$ , and after calculation similar to the arguments before, we have the weights as

$$w_2 = \frac{N/I + 1 - e^{j\Delta\theta}}{N/I + 1 - e^{-j\Delta\theta}} w_1$$

The output SINR becomes

$$\text{SINR} = \frac{S \cos^2 \varphi}{I \cos^2(\Delta\theta / 2 + \varphi) + N/2} \quad (\text{MMSE receiver})$$

where  $\varphi = \tan^{-1} \frac{\sin(\Delta\theta)}{1 - \cos(\Delta\theta) + N/I}$ .

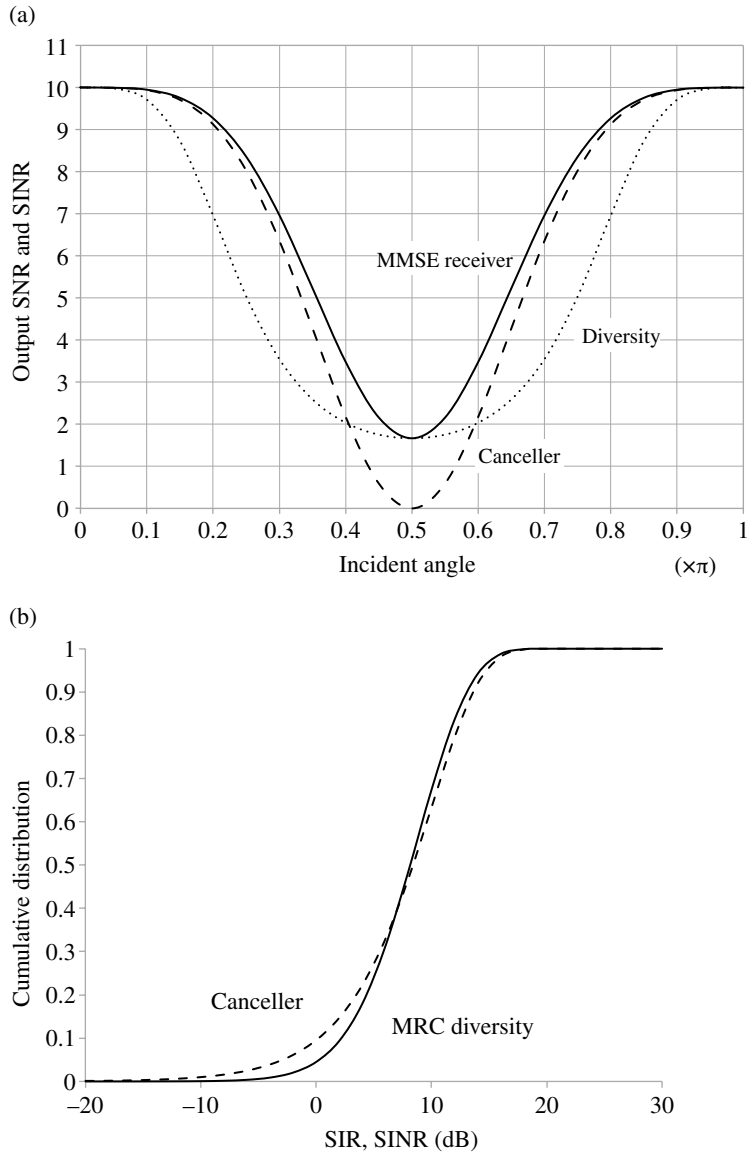
The same result is obtained by maximizing the following SINR expression with respect to  $w_1$  and  $w_2$ :

$$\text{SINR} = \frac{|w_1 + w_2|^2 S}{|w_1 + e^{-j\Delta\theta} w_2|^2 I + (|w_1|^2 + |w_2|^2) N}$$

A numerical example is shown in Figure 7.57. The MMSE receiver shows the best performance all over the incident angle.

We notice that the antenna weights  $w_1$  and  $w_2$  take the same absolute value (phased array system). This is due to the assumption of plane wave propagation of signals.

*Performance Example II.* We consider an adaptive array antenna interference cancelling system with two antenna elements under independent Rayleigh fading environments. Under the Rayleigh fading condition, a signal is conveyed with many plane waves having different propagation paths, and the received signal powers become



**FIGURE 7.57** Interference cancelling receiver performance ( $d = \lambda/2$ ). (a) Receiver performance (input SNR=5, SIR=2) and (b) cumulative distribution of S(I)NR under Rayleigh fading. Average SNR ( $b_1/N$ ) = 10 dB, SIR ( $b_1/b_2$ ) = 7 dB.

different at the two antenna elements. And therefore the diversity effect appears and the optimum antenna weights take different absolute values in contrast to example I. A desired signal and one interfering signal are assumed. The performance of the system is given only statistically since the Rayleigh fading is a stochastic process.

As the performance measure, the cumulative distribution of SINR is investigated. We cannot obtain theoretically the performance for the MMSE receiver in this situation. Instead of the MMSE receiver, the MRC diversity system and interference cancelling receiver are considered. The cumulative distribution with the MMSE receiver is over-estimated as the best one of the two systems in all the range of SINR values as shown in the performance example I.

*Interference Canceller.* The two antenna weights are controlled to cancel the interference signal at any time. Then the diversity effect is lost to cancel completely the interference. Then the performance of the system is measured with signal to noise power ratio.

Probability density function of received signal power under the Rayleigh fading is given as (Section 4.3)

$$p_R(q) = \frac{1}{b} e^{-q/b}$$

where  $b$  is the average power ( $b = \langle q \rangle$ ). The distribution function of received power,  $\text{Prob}(q \leq q_0) \equiv P_R(q_0)$ , becomes

$$\begin{aligned} P_R(q_0) &= \int_0^{q_0} p_R(r) dr \\ &= 1 - e^{-q_0/b} (1 + q_0/b) \end{aligned}$$

SNR distribution of the canceller is given by replacing  $q_0$  by  $qN$ , where  $q$  is the SNR value and  $N$  is the noise power.

*MRC Diversity Receiver.* The two weights are controlled to get MRC diversity without taking the interference into consideration. The probability density function of desired signal power becomes (Eq. 7.3a,  $M = 2$ )

$$p_R(q) = \frac{q}{b^2} e^{-q/b}$$

The interference signal power is subjected to the exponential distribution (without diversity combining). Therefore, the cumulative distribution function,  $Pr(q) = \text{Prob}(\text{SINR} \leq q)$ , of  $\text{SINR}(q = \gamma_1 / (\gamma_2 + N))$  of the system is given as

$$P_R(q) = \int_0^{q(\gamma_2 + N)} d\gamma_1 \int_0^\infty p_1(\gamma_1) p_2(\gamma_2) d\gamma_2$$

where  $p_1(\gamma_2)$  and  $p_2(\gamma_2)$  are the probability density functions of received power for desired and interfering signals, respectively. For this system, performing the integration, we get

$$P_R(q) = 1 - \left[ \frac{qN/b_2 + b_1/b_2}{q + b_1/b_2} + \frac{qb_1/b_2}{(q + b_1/b_2)^2} \right] e^{-qN/b_1}$$

where  $b_1$  and  $b_2$  and  $N$  denote average power per antenna for the desired and interfering signals and noise, respectively.

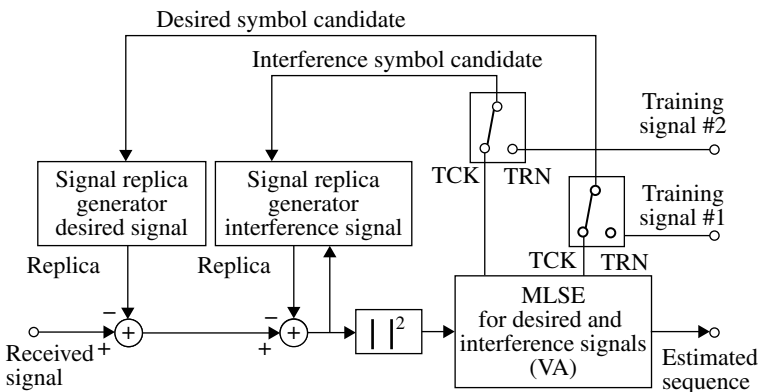
The cumulative distribution of SNR for the canceller and SINR for the MRC receiver is shown in Figure 7.57b.

### 7.6.2 Adaptive Interference Suppression

The antenna array system cannot cancel an interfering signal without cancelling the desired signal, if it is in the same direction as the desired signal. The other category of adaptive interference cancellation can cope with this case. Figure 7.58 shows an adaptive interference canceller based on MLSE without an array antenna [82, 83]. The MLSE or the Viterbi algorithm is used for estimating both the desired and the interfering signal. Adaptive equalizers are also assumed in this system. The adaptation algorithm used in the equalizers is the recursive least squares method (Section 2.5.2). The reference signals for the desired and interfering signals are produced, using the candidates for desired and interfering signals and the channel impulse response. They are subtracted from the received signal to output an error signal. The squared error signal is the target function for the MLSE. Training signals are assumed for initial acquisition. The adaptive interference canceller/equalizer is further combined with a diversity receiving system. Good results are obtained with computer simulation. Trellis-coded modulation is introduced to this adaptive interference canceller in [84]. Computer simulation shows better results than that without trellis-coded modulation.

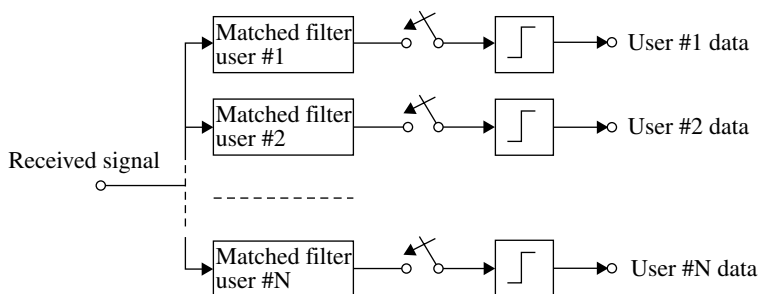
For suppressing the effect of the interfering signal, FEC using information about interference level is proposed [85]. The interference signal is periodically measured in the absence of signal transmission.

Adaptive interference cancellation is discussed in the following for spread-spectrum code division multiple-access (SS-CDMA) systems. A primitive multiuser detector for a CDMA system consists of a bank of matched filters and a threshold device as shown in Figure 7.59. The performance of the receiver suffers from the cross



**FIGURE 7.58** Adaptive interference canceller using RLS-MLSE [82]. Reproduced by permission of © 1994, IEICE.



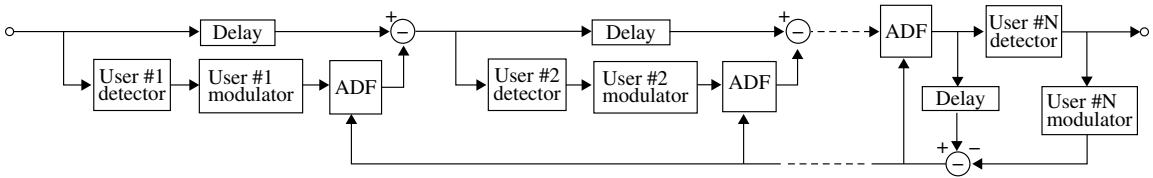


**FIGURE 7.59** A multiuser detector for a CDMA system.

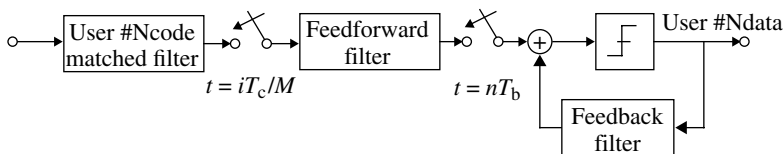
correlation between the spectrum-spreading codes for different users. The optimum multiuser receiver can be obtained with the maximum likelihood sequence detector. The complexity of the receiver, however, increases exponentially with the number of users. Therefore, some suboptimum receivers are proposed to reduce the receiver complexity [86–90]. A multistage adaptive interference canceller is shown in Figure 7.60. The interference from users #1 through  $\#(N-1)$  is subtracted successively from the input signal to get the desired signal (user  $\#N$ ). Tentative decisions on the data are used to produce replica signals of the interfering signals. Knowledge about the other users' spectrum-spreading codes is assumed in this receiver. Adaptive interference cancelling receivers for SS-CDMA systems (Fig. 7.61) are discussed in [91] and [92] for systems where other users' codes are unknown. The received signal is fed to a filter matched to the desired (user  $\#N$ ) signal SS code. The output of the matched filter is sampled at a fractional time interval of the chip rate. The interference from other users is suppressed with the adaptive feed-forward filter using the orthogonal property of the codes. The feedback filter equalizes the intersymbol interference. Any algorithm such as the MMSE method can be used to adapt the filter. Improvements in adaptation speed in fading channels are reported in [84] and [93].

### 7.6.3 Discussion

Some people would consider that antenna arrays are different from diversity systems. However, as seen from Figures 7.2 and 7.54, an antenna array has the same structure as a cophased diversity combining system. If we consider a diversity combining system, similar to an antenna array, the diversity system gives a directivity pattern that is determined from the weighting factor. (The directivity pattern for the receiver is the same as for a transmitter, because of the reciprocity of a linear circuit.) The difference between an antenna array and a diversity combining system is only superficial and consists in the assumption of arrival signals. Array antennas usually assume a desired signal and a small number of interfering signals, all of which are plane waves; on the other hand, diversity antennas assume a desired signal that is a sum of many plane wave signals, which are independent of each other. Consequently, the correlation of received signals between antennas is high for an antenna array and low for a diversity system. These situations correspond to environments for mobile terminals and base stations.



**FIGURE 7.60** A multistage adaptive interference cancelling receiver for an SS-CDMA system. ADF denotes adaptive digital filter [87]. Reproduced by permission of © 1990, IEEE.



**FIGURE 7.61** Interference cancelling SS-CDMA receiver without knowledge of other users' codes.  $T_c$  and  $T_b$  are the chip and data duration, respectively.  $M$  is an integer.

Using a proper combining algorithm such as the LMS error algorithm, an antenna array and a diversity system work in the same manner. For example, consider a situation where an echo signal with a time delay arrives at a receiver in a direction different from the main signal; then the echo signal may be used or cancelled depending on the time delay in both the antenna array and the diversity system.

## 7.7 VOICE CODING

The most straightforward system for voice coding [56, 94] is the usual analog-to-digital conversion at periodic sampling instants. In this system, the sampled signals are quantized into discrete levels spaced at a given distance. The difference between the input signal level and the quantized level is called quantization error or quantization noise. The quantization noise power is calculated as follows. Quantization noise is assumed to be independent and uniformly distributed in the range of  $-\Delta/2$  to  $\Delta/2$ , within the frequency range  $-f_s/2$  to  $f_s/2$ , where  $\Delta$  is the level spacing and  $f_s$  is the sampling frequency. The expected value of the noise power  $N$  is then

$$N = \int_{-\Delta/2}^{\Delta/2} x^2 p(x) dx$$

$$= \frac{\Delta^2}{12}$$

where the probability density function  $p(x) = 1/\Delta$  for  $|x| \leq \Delta/2$  and  $p(x) = 0$  for  $|x| \geq \Delta/2$  is used. Let us consider the case of quantizing a sinusoidal signal of peak value  $A_p$ . Using  $\Delta$ , we have the average power  $S = \frac{1}{2} A_p^2 = \frac{1}{2} (2^{n-1} \Delta)^2$ . The average signal to quantization noise power ratio becomes

$$\frac{S}{N} = \frac{\frac{1}{2} \cdot 2^{2(n-1)} \Delta^2}{\frac{\Delta^2}{12}}$$

$$= \frac{3}{2} \times 2^{2n}$$

$$\approx 6n + 1.8 \text{ [dB]}$$

### 7.7.1 Pulse Code Modulation

When we decrease the signal level in the above coding system, the signal to noise power ratio becomes small for a fixed  $n$  and  $\Delta$ . This scheme is not appropriate for voice signal transmission, since the voice signal level is not always at a constant value. To solve this problem, the pulse code modulation (PCM) system employs a technique that compresses the input signal to the quantizer as shown schematically in Figure 7.62. The input signal is compressed instantaneously depending on its level. The effect of compressing is equivalent to decreasing  $\Delta$  for lower-level signals. Thus, the signal to noise power ratio becomes less dependent on the input signal level. The signal is expanded at the receiver to recover the original, uncompressed signal.

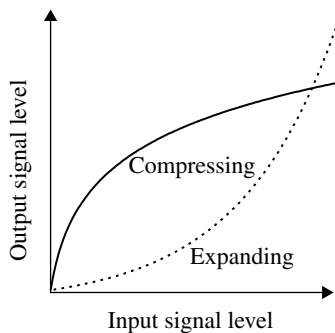
The CCITT has standardized two compression laws [56]: the  $\mu$ -law

$$y = \frac{\text{sgn}(x)}{\ln(1 + \mu)} \ln \left( 1 + \mu \left| \frac{x}{x_p} \right| \right) \quad \left( \left| \frac{x}{x_p} \right| \leq 1 \right)$$

and the A-law

$$y = \begin{cases} \frac{A}{1 + \ln A} \frac{x}{x_p} & \left( \left| \frac{x}{x_p} \right| \leq \frac{1}{A} \right) \\ \frac{\text{sgn}(x)}{1 + \ln A} \left[ 1 + \ln A \left| \frac{x}{x_p} \right| \right] & \left( \frac{1}{A} \leq \left| \frac{x}{x_p} \right| \leq 1 \right) \end{cases}$$

where  $x_p$  is the maximum input signal level and  $\mu = 100$  and  $255$  and  $A = 87.6$  are the standardized values. The sampling frequency is  $8 \text{ kHz}$  and there are  $2^8$  quantization levels. The voice coding rate is then  $64 \text{ kbps}$ . PCM is seldom used for mobile radio communications because its transmission rate is high, requiring a wide bandwidth.



**FIGURE 7.62** Compression and expansion of voice signal.

7.7.2 Delta Modulation

The coder and decoder for a delta modulation system are shown in Figure 7.63. The local signal  $s_l(t)$ , which is generated by using previously decoded signal, is subtracted from the input signal  $s_i(t)$  to produce an error signal  $e(t)$ . The polarity of the error signal is detected, and a “1” or “0” signal is assigned as the coded output signal (i.e., one-bit quantization). The decoder is the same as the feedback circuit, called a local decoder, in the encoder.

The local decoder tracks the input signal by adding a step size of  $\Delta$  or  $-\Delta$  to its value, that is, integrating at every clock time (Figure 7.63). The integrator in the local decoder is actually replaced with a more general low-pass filter. The filter may be called a predictor, which predicts the next signal from the previous signals. The minimum error in tracking is  $\pm\Delta$ , and using a smaller  $\Delta$  corresponds to a higher signal to noise ratio. However, tracking speeds become slower for a smaller  $\Delta$ , and a larger error is produced for rapidly changing input signals, as shown in Figure 7.64. The tracking speed is raised by increasing the coding clock frequency.

The adaptive delta modulation (ADM) system changes its step size in order to track the rapidly changing input signal without increasing the coding frequency.

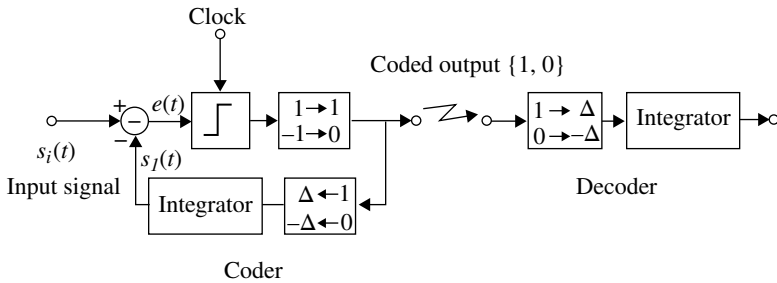


FIGURE 7.63 Delta modulation system.

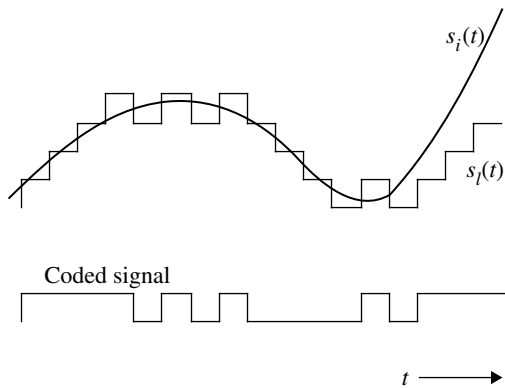
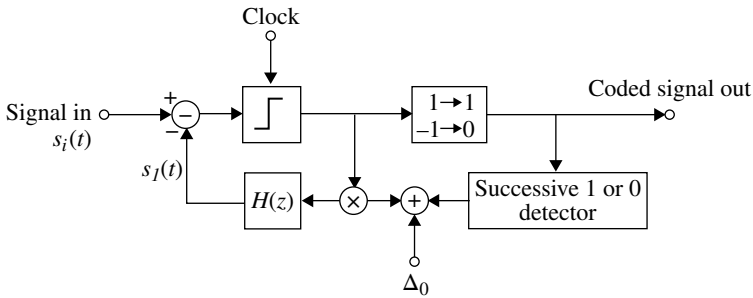


FIGURE 7.64 Waveforms for delta modulation system.



**FIGURE 7.65** Adaptive delta modulation encoder.

The block diagram of the ADM encoder is shown in Figure 7.65. A general filter is assumed for the local decoding filter. If a successive 1 or 0 detector detects a given number of successive occurrences of 1's or 0's, it outputs a value to be added to the given step size  $\Delta_0$ , increasing the total step size. The decoder at the receiver is the same as the local decoder of the encoding circuit. The receiver decoder outputs the replica signal  $s_f(t)$  (assuming no transmission errors) using the received data signal and knowing  $H(z)$ ,  $\Delta_0$ , and the structure of the successive 1 or 0 detector. ADM is used for mobile radio communications at a coding speed of 10–30 kbps.

### 7.7.3 Adaptive Differential Pulse Code Modulation

The adaptive differential pulse code modulation (ADPCM) system uses a multilevel quantizer and an adaptive prediction filter. Figure 7.66 shows the ADPCM system standardized by the CCITT. The sampling frequency is 8 kHz, and there are  $2^4$  quantization levels, resulting in a coding rate of 32 kbps.

The quantizer and predictor are shown in more detail in Figure 7.67. If the quantization error,  $e(n) - e'(n)$ , is small, the signal  $\tilde{x}(n) = \hat{x}(n) + e'(n)$ , which is obtained at the receiver output, is close to the input signal  $x(n)$  so long as there are no transmission errors. In Figure 7.66, if we assume analog transmission by removing the quantizer, encoder, and decoder and connecting their inputs and outputs, exactly the same signal as  $x(n)$  is obtained at the receiver output. This means that the transfer function of the transmit circuit is the inverse transfer function of the receiver circuit. The receive feedback filter is called the synthesis filter, which is also used in the feedback path of the transmit filter. The number of bits for quantization can be decreased, since, due to prediction, the dynamic range of  $e(n)$  is reduced compared with that of  $x(n)$ .

Since the adaptation is done using the variable  $e'(n)$ , which can be obtained from the transmitted signal sequence  $c(n)$ , we do not need to send any extra information to control the receiver adaptive filter.

The 32 kbps ADPCM is adopted for the European and Japanese digital cordless telephone system (Section 9.6).

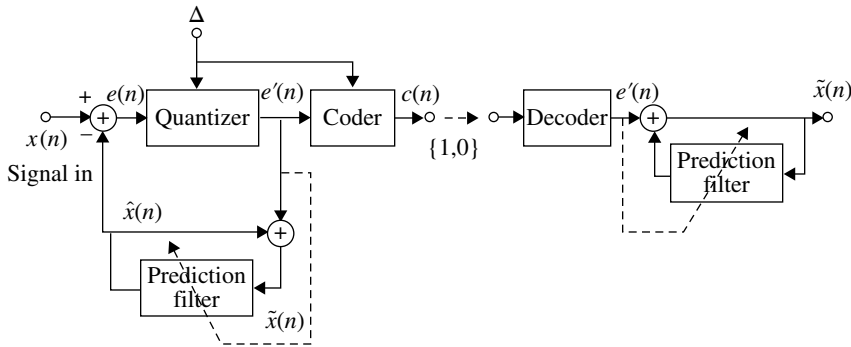


FIGURE 7.66 Adaptive differential PCM system.

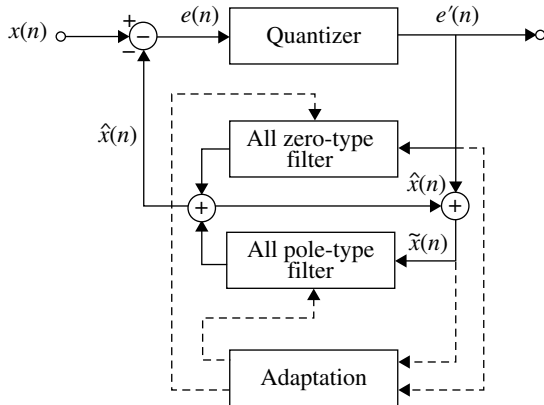
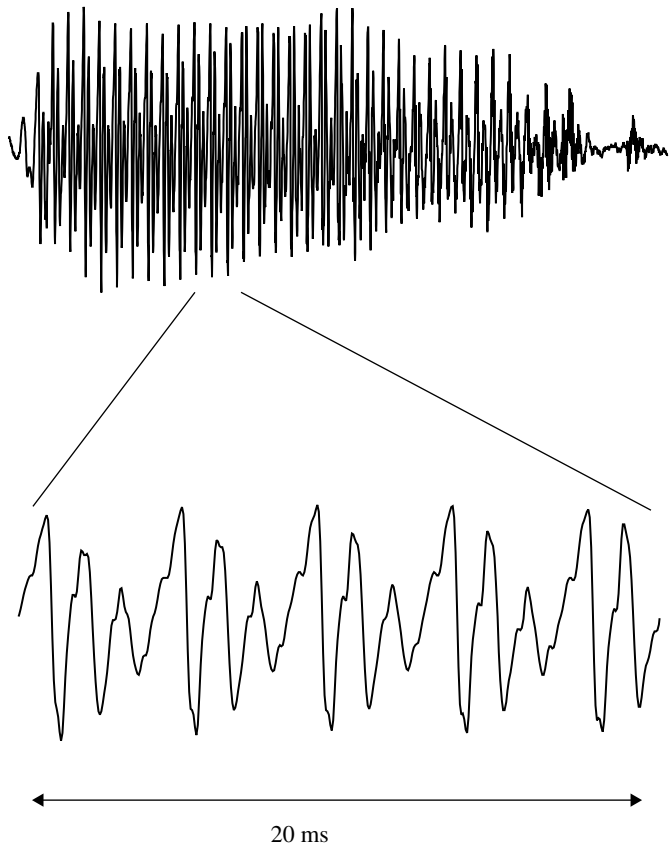


FIGURE 7.67 Detailed block diagram of ADPCM predictor.

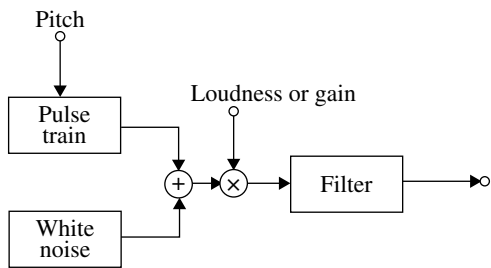
7.7.4 Adaptive Predictive Coding

If more sophisticated prediction is used, such as adaptive predictive coding (APC), coding efficiency can be increased. This scheme is based on the periodic structure of voice signals. An example of a voice signal waveform is shown in Figure 7.68. It is the Japanese pronunciation of SAI by a female college student. We can see some regularities in the waveform due to the mechanism of voice signal generation. A voice signal generation model is shown in Figure 7.69. Periodic pulses and white noise are used as the sound sources. The former are generated as to model the vibration of vocal cords to produce voiced sound. The vibration frequency, or pitch, corresponds to the repetition of peak pulses in Figure 7.68. The white noise is generated for the fricative (or unvoiced) sounds. The filter represents the vocal tract from the sound source input to the output, in other words, the mouth, lips, and nose.

The regularity of the waveform does not change for several tens of milliseconds. Thus, the regular waveform in some time period can be expressed by a small set of parameters, such as the pitch, loudness, and transfer function of the vocal tract, or coefficients of the equivalent short-term prediction filter.



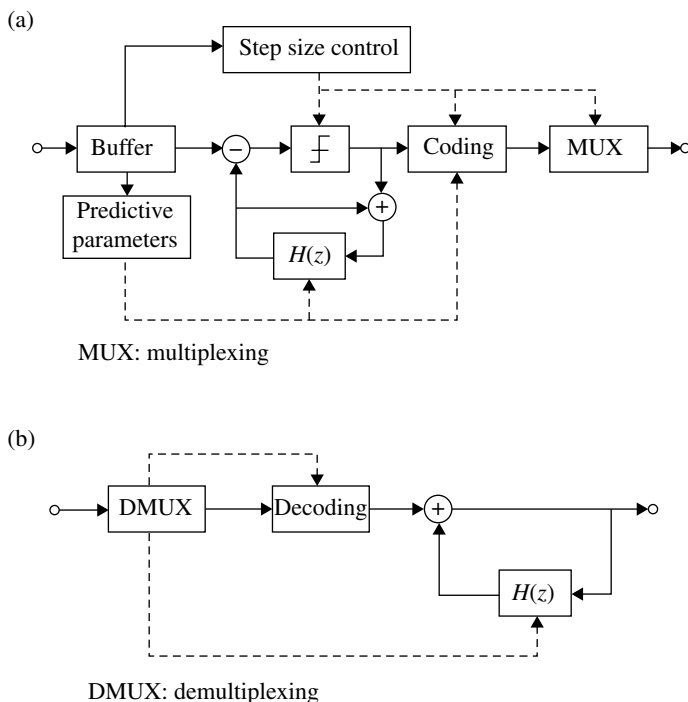
**FIGURE 7.68** Example of a voice signal waveform.



**FIGURE 7.69** Model for generation of a voice signal.

The coefficients of the short-term prediction filter are obtained with linear predictive coding (LPC) analysis. In short, the LPC analysis determines the transfer function of the vocal tract using the short-term voice signal as a test signal. A more detailed description of a LPC analysis is beyond the scope of this book; refer [56, 94–96] for further information. APC is used for most of the efficient voice coding systems.





**FIGURE 7.70** Adaptive predictive coding system. (a) Coder and (b) decoder.

The long-term or pitch prediction is made so as to minimize the squared pitch prediction error. Assume a first-order digital filter; then the error is defined as

$$E = \sum_{n=0}^{N-1} [e(n) - \beta e(n-T)]^2$$

where  $e(n)$  is the short-term prediction error given by LPC,  $\beta$  is a coefficient,  $T$  is the pitch duration, and  $N$  is the number of samples of the signal in a block.  $\beta$  and  $T$  are found so as to minimize  $E$ .

Figure 7.70 shows an APC system. The major difference of this coding from the ADPCM system consists in the way of estimation and transmission of the adaptive parameters. In the APC system, the input signal is stored in a buffer for a time period of, say, 20 ms; then the short-term parameters are analyzed by LPC, followed by long-term estimation. The quantization step size is also determined using the buffered signal. These parameters, the side information, are multiplexed and transmitted with the coded signal.

The parameter estimation and coding are carried out independently in the systems so far described. In the following systems, sometimes called hybrid coding systems, combined analysis of the parameter estimation and coding is made using a cost function for the minimum waveform error between input signal and locally decoded signal.

### 7.7.5 Multipulse Coding

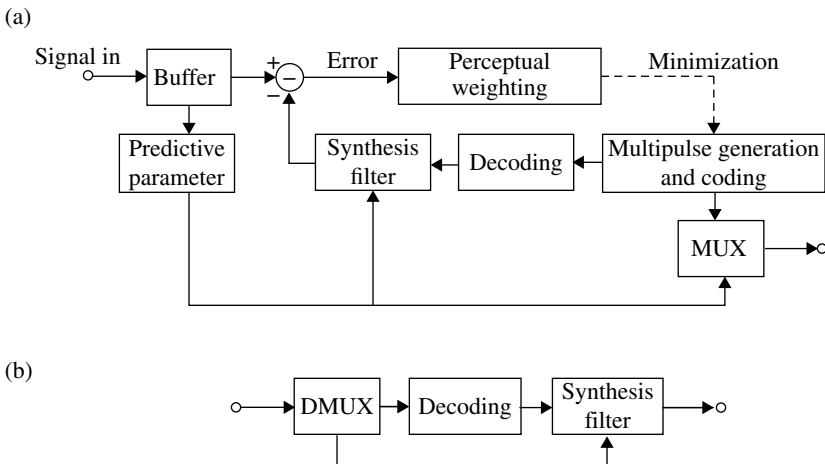
Figure 7.71 shows a block diagram of the multipulse coding system. The estimation of short-term and/or long-term parameters is made using the buffered signal for a given size of block. For the sound source input to the synthesis filter, many pulses with different amplitude and position are generated to achieve the minimum cost function:

$$E = \sum_{n=0}^{N-1} \left[ \left\{ x(n) - \sum_{j=1}^k g_j h(n - m_j) \right\} * w(n) \right]^2$$

where  $x(n)$  are samples of the input signal,  $g_j$  and  $m_j$  are the amplitude and position of  $j$ th pulse,  $h(n)$  are samples of the impulse response of the synthesis filter, and  $w(n)$  are samples of the impulse response of the perceptual weighting factor. The symbol  $*$  denotes convolution. The perceptual weighting factor is used to improve voice quality as heard by the human ear. The coding (analysis) is carried out by feeding back the locally decoded (synthesis) signal and is called the analysis-by-synthesis (A-b-S) method. The key technical issue in multipulse coding is determining the multipulse and its efficient coding for good voice quality. Multipulse coding requires a high volume of signal processing. To reduce the amount of signal processing, intervals of pulses are constrained to given values (regular pulses) in the pan-European (GSM) digital cellular speech coding: regular pulse excited LPC with long-term prediction (RPE-LPC/LTP) [97].

First, the LPC predicted error is calculated as

$$e(n) = x(n) - \sum_{i=1}^p a_i x(n-i)$$



**FIGURE 7.71** Multipulse coding system. (a) Coder and (b) Decoder.

where  $a_i$  and  $P$  are the coefficients and order of the LPC filter. The pitch parameters are given by minimizing

$$E(\beta, M) = \sum_{n=0}^{N-1} [e(n) - \beta v(n-M)]^2$$

where  $N$  is the subframe length (say, 5 ms, i.e.,  $N=40$ ),  $\beta$  is the pitch coefficient,  $M$  is the pitch duration, and  $v(n)$  is past subframe sound source expressed with the regular pulses. The regular pulses  $x_m(i)$  are made by sampling  $d(n)$ , defined by  $e(n) - \beta v(n-M)$ , at some interval, say, 4,

$$x_m(i) = d(m + 4 \cdot i) \quad (m=0,1,2,3)$$

where  $m$  denotes the initial phase of a regular pulse. The initial phase is determined by  $\max \sum_{i=0}^{N_{\text{sub}}} x_m^2(i)$ , where  $N_{\text{sub}}$  (say, 40) is the subframe length. The amplitudes of the regular pulses are coded with adaptive PCM.

The transmitted information comprises the LPC predictive parameters for each frame length (20 ms): the pitch duration  $M$ , the pitch coefficient  $\beta$ , and the amplitudes and the initial phase for each subframe length (40 samples = 5 ms).

### 7.7.6 Code-Excited Linear Predictive (CELP) Coding

Code-excited linear predictive (CELP) coding is different from multipulse coding because it uses prepared waveforms instead of multipulse waveforms as the source signal input to the synthesis filter (Fig. 7.72). The selection of one of the waveforms stored in the codebook and the gain factor is made to minimize the perceptually weighted error power. The index for a waveform and the gain factor are transmitted together with parameters of the synthesis filter. Assigning an index to a waveform of a given number of samples corresponds to vector quantization. Using the vector quantization, the average number of bits per sample may become less than 1.

The long-term or pitch prediction filter in the synthesis filter can be adaptively controlled with the error signal. For this case, the synthesis filter is given in Figure 7.73. The pitch predictor consists of a first-order filter. The delay and coefficient  $\beta$  are adjusted to minimize the perceptually weighted error power. A subframe data block of, say, 5 ms is stored and delayed in the shift register. The register is called the adaptive codebook.

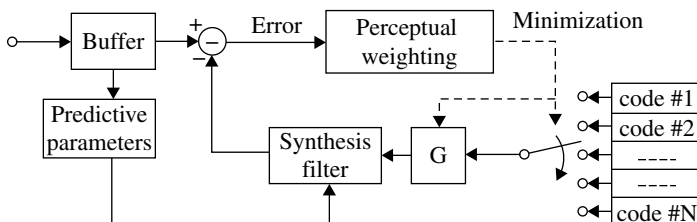
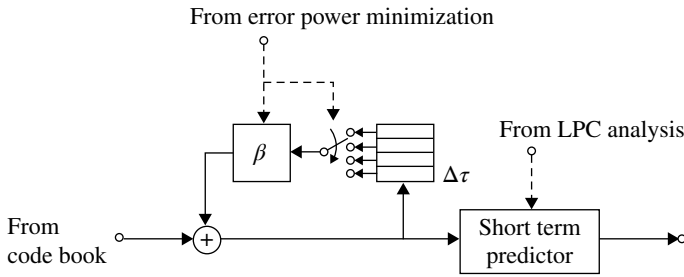
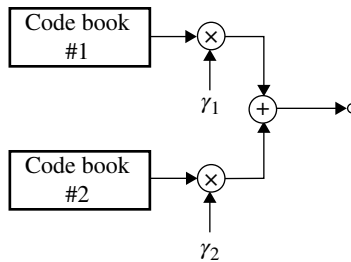


FIGURE 7.72 Code-excited linear predictive (CELP) coding.



**FIGURE 7.73** CELP with adaptive codebook (pitch prediction).



**FIGURE 7.74** Sound source codebook for VSEL.

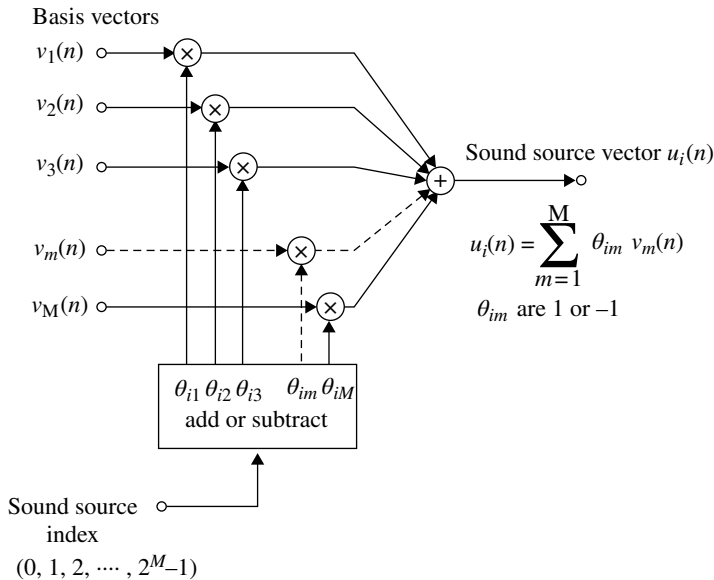
The vector sum excited LP (VSEL) coding is proposed to decrease the computational complexity for the optimum code search and memory for the codebook. In addition, it has immunity against transmission errors. Multiple codebooks are used for the sound source in the VSEL as shown in Figure 7.74. Each codebook is implemented as shown in Figure 7.75. Consider that the number of sound source is  $2^M$  (indexed by  $i$  as  $i=0,1,2, \dots, 2^M-1$ ). We prepare  $M$  different basis vectors, which are determined by experiment and standardized. A sound source vector is generated as

$$u_i(n) = \sum_{m=1}^M \theta_{im} y_m(n) \quad i=0,1,2,\dots,2^M-1$$

where  $\theta_{im}$  takes 1 or  $-1$  corresponding to sound source index  $i$ . Thus,  $2^M$  sound sources are produced with stored basis vectors. The index of  $M$  bits for each codebook is transmitted at each subframe. The receiver constructs a sound source with the above equation and the circuit of Figure 7.71b. The transmitted analog signal is obtained by inputting the sound source to the synthesis filter (Fig. 7.72).

Consider that a transmission error occurs in the  $M$  bits: the polarity of one of the  $M$  basis vectors is reversed. The generated sound is not much different from the original sound. This explains why VSEL has immunity against transmission errors.

The use of two codebooks helps to reduce the amount of signal processing required. Description of this topic is beyond the scope of this book.



**FIGURE 7.75** Implementation of sound source generator for VSELP.

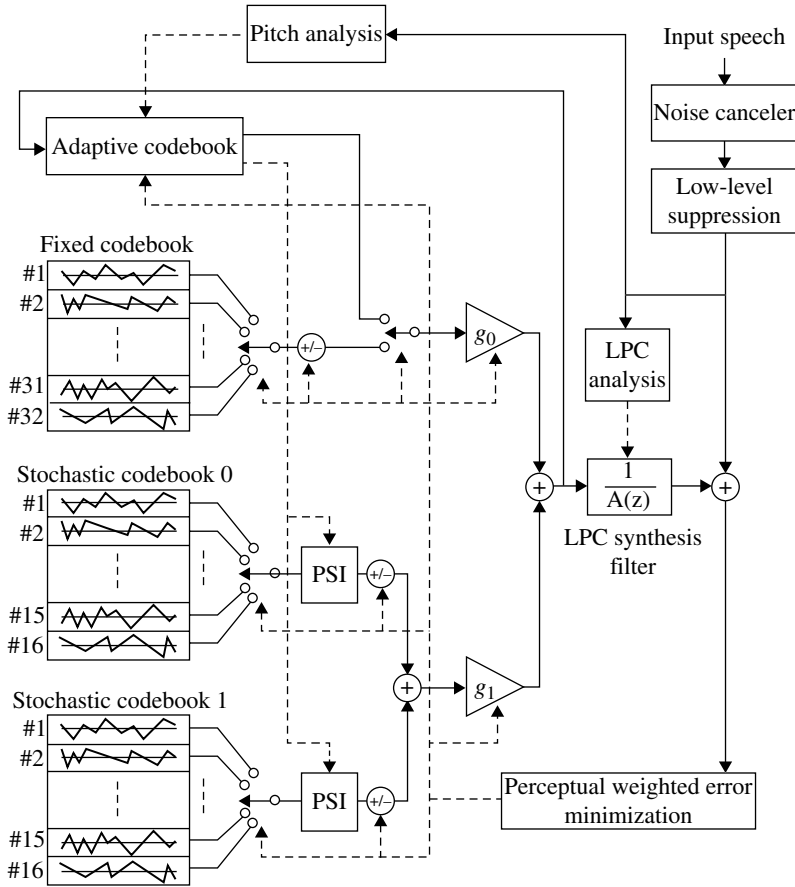
VSELP was selected as the full rate standard, of which the coding rate is 13 kbps (7.95 kbps for voice coding and the rest for error control) and 11.2 kbps (6.7 kbps for voice coding and the rest for error control) for American and Japanese digital cellular telephone systems, respectively [98, 99].

For the half-rate standard in Japanese digital cellular telephone systems, pitch synchronous innovation CELP (PSI-CELP) [100] was adopted. The coding rate is 5.6 kbps with 3.45 kbps for voice coding and 2.15 kbps for error control. The PSI-CELP coder is shown in Figure 7.76. The lengths of the frame and subframe are 40 and 10 ms, respectively. The pitch prediction, predicted error, and codebook gain are coded in a subframe. It has three kinds of codebooks for quantization of source signal: adaptive (for pitch prediction), fixed, and the stochastic codebooks. The adaptive codebook and the fixed codebook are selected for voiced and unvoiced signals, respectively. The most significant point of the PSI-CELP is that the stochastic codes are synchronized to the pitch.

**ACELP:** In the 1990s, performance of CELP was improved. Algebraic CELP (ACELP) is known as a representative one. This coding with a coding rate of 8 kbps shows voice quality equivalent to 32 kbps ADPCM. The contents of codebooks are calculated with an algebraic operation on pulses, instead of a conventional waveform with vector quantization. For example, G.729 (conjugate structured ACELP (CS-ACELP)), which was standardized in ITUT-T in 1995, has codebooks of which contents are given as

$$c[n] = S_0 \delta[n - m_0] + S_1 \delta[n - m_1] + S_2 \delta[n - m_2] + S_3 \delta[n - m_3]$$

$$(n = 0, 1, \dots, 39)$$



**FIGURE 7.76** Pitch synchronous innovation CELP (PSI-CELP) coder [100]. Reproduced by permission of © 1993, IEICE.

where  $\delta[n]$  is a unit pulse, that is,

$$\delta[n] = \begin{cases} 1 & (n = 0) \\ 0 & (n \neq 0) \end{cases}$$

and  $S_j (j = 0, 1, \dots, 3)$  takes  $\pm 1$ .

Thus, the codebook contents consist of four pulses, which take an amplitude of  $\pm 1$  and therefore, we can call this as a multipulse coding. The pulse time positions are given by  $m_0, m_1, m_2$ , and  $m_3$ . In G.729, the four positions are selected as the best one from candidates, which are listed in Table 7.3. To express a codebook, each of  $m_0, m_1, m_2$  requires 3 bits and  $m_3$  needs 4 bits. One bit is required for amplitude of the pulses, and consequently a codebook consists of 17 bits. A codebook is selected in every subframe of 40 samples. The sampling frequency is 8 kHz, and the subframe

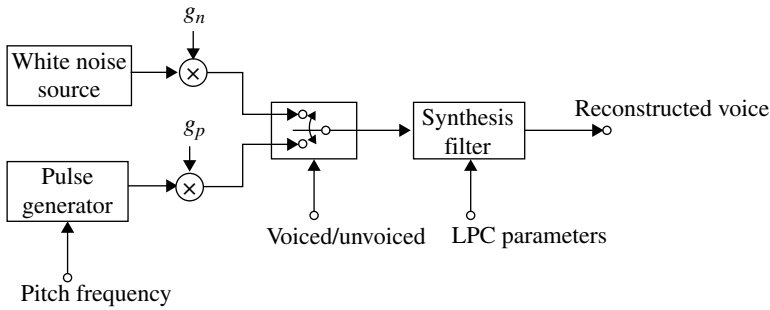
**TABLE 7.3 Pulse Position Candidates in ACELP**

Pulse number	Pulse position candidates
$m_0$	0,5,10,15,20,25,30,35
$m_1$	1,6,11,16,21,26,31,36
$m_2$	2,7,12,17,22,27,32,37
$m_3$	3,8,13,18,23,28,33,38
	4,9,14,19,24,29,34,39

duration becomes 5 ms. The transmission speed for the content of a codebook is 3.4kbps. Since the total coding rate of G. 729 is 8kbps, the remaining 4.6kbps is assigned for transmission of other information including filter parameters.

Coding bits are reduced owing to one-bit expression of the pulse amplitudes. Computational complexity for analyses and synthesis is also decreased due to the same reason. Required memory size is reduced since exciting vectors are produced by calculation. A CS-CELP encoder in G.729 consists fundamentally of fixed codebook (Fig. 7.72) and adaptive codebook (Fig. 7.73). ACELP is intended to be used for the future public land mobile telecommunication systems (FPLMTS). Therefore, schemes that show immunity against transmission errors are adopted. One of the schemes is gain codebook, which has a conjugate structure: gains for a pitch period adaptive codebook and for algebraic noise codebook are vector quantized as a whole and stored in vector quantized gain codebook. The gain codebook is expressed with a sum of two conjugate structured subcodebooks. Then even if an index of a subcode is erroneously received, the effect of the error on recovered voice quality is suppressed as long as the index for the other subcodebook is correct. Furthermore, the memory size of codebook as a whole is reduced owing to the conjugate structure. Dynamic range of the gain for the noise codebook is reduced by introducing a backward expectation technique. LPC coefficients are not directly quantized and transmitted: the coefficients are transformed into the line spectral pair (LSP) coefficients, and then quantization of the LSP coefficients and linear interpolation between the subframes are performed to raise coding efficiency.

ACELP were introduced to cellular systems to enhance their voice quality [101]. In the GSM system, it is standardized in 1996 as enhanced full rate (EFR) with a coding rate of 22.8 kbps, which is the same as the conventional coding method (RPE-LTP). In 1998, it is standardized as adaptive multirate (AMR), which adaptively takes a transmission speed of 4.75, 5.15, 6.7, 7.95, 10.2, and 12.2kbps. This is approved by 3GPP for IMT-2000 in 1999. In the American digital cellular D-AMPS system, the EFR (13kbps) was standardized in addition to the conventional VCELP in 1996. For cdmaOne (3GPP2) system, enhanced variable bit rate coder (EVRC) (IS-127) is introduced in addition to the conventional Qualcomm CELP (QCELP) in 1997. The transmission speeds are 8, 4, and 0.4kbps. In personal digital cellular (PDC) system, ACELP with transmission speed of 11.2kbps was standardized in addition to the conventional VCELP and PSI-CELP in 1998. This is CS-ACELP with countermeasures against bit errors in mobile radio communication [102].



**FIGURE 7.77** Decoder for LPC vocoder.

### 7.7.7 LPC Vocoder

In general, the voice decoders reconstruct the transmit voice signal by applying the source signal to the (synthesis) filter. Similarly to the other LPC schemes, the LPC vocoder uses the LPC synthesis filter. In this system, the input signal to the synthesis filter is simplified as shown in Figure 7.77.

At the encoder, the LPC short-term analysis, the pitch analysis, and voiced or unvoiced signal decision are carried out. The unvoiced signal and voiced signal are represented by white noise and a periodic pulse train, respectively. The pitch frequency, voiced or unvoiced signal indicator, the gain factor  $g_n$  or  $g_p$ , and synthesis filter parameters are transmitted. Since the coding rate is as low as 2.4 kbps, the voice quality is not sufficient for use in public communications at the time of writing.

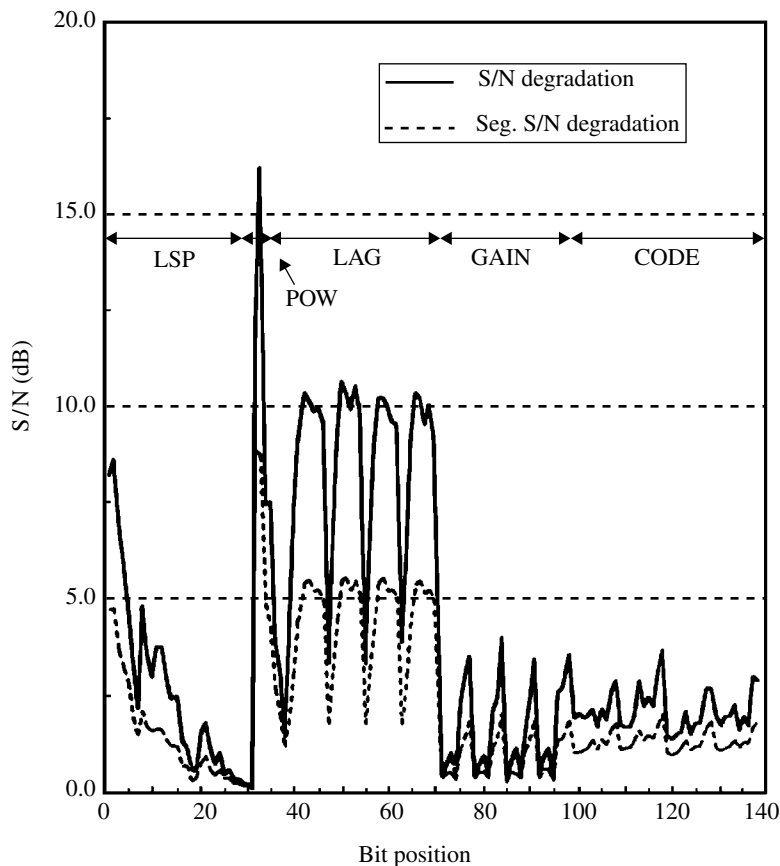
### 7.7.8 Application to Mobile Radio Communications

When we apply voice coding to mobile radio communications, the voice quality, coding rate, coding/decoding delay, and required signal processing volume become important issues. The voice quality must be considered with as well as without transmission errors. The coding rate affects the spectrum efficiency and its importance depends on the system. Voice coding methods currently used in public mobile radio communication ranges from 32 kbps ADPCM for digital cordless telephones to 5.6 kbps PSI-CELP for the Japanese half-rate digital cellular system. When coding/decoding delays become long, echo suppression/cancellation must be introduced. The signal processing requirement is being reduced by virtue of improvements in microsemiconductor technologies.

Burst error as well as random error in mobile radio channels causes degradation in decoded voice quality. The degree of degradation depends on the voice coding scheme. (Adaptive) Delta modulation is immune to transmission error: the acceptable average bit error rates are around  $10^{-2}$  for a relatively low coding rate, say, of 16 kbps.

The transmission errors produce annoying click noise in PCM or ADPCM systems. Techniques have been proposed to reduce the effects of click noise [103–105]. Two kinds of techniques are known: noise blanking and replacement. In the former





**FIGURE 7.78** Transmission error effect on voice quality for the PSI-CELP [100]. Reproduced by permission of © 1993, IEICE.

method, the signal is killed during the click noise period. In the latter method, the noisy signal period is replaced by the previous period. In these techniques, smooth continuation of the voice signal is favorable. The click noise is detected using error detection coding and/or by differentiating the signal, detecting abrupt change in the signal level. For ADPCM, a technique is proposed in which correlation between the coefficients of the prediction filter for voice signal is used: when an error is detected, the coefficients are replaced with those of the previous signal period. For a highly efficient (i.e., low coding rate) coding system, transmission errors cause intolerable deterioration in voice quality. In this system, error control is a must. The impact of errors on the voice quality depends strongly on the bits assigned for the synthesis filter parameters and sound source. For example, Figure 7.78 shows how the signal to noise power ratio depends on errors in the different bits assigned for PSI-CELP. The delay parameters used for the pitch prediction and the pitch synchronization are fairly sensitive to transmission errors. To protect the bits with different sensitivity

from transmission errors, error correction codes with different degree of error correcting capability for respective information-bearing bits are used [106].

For a long burst error resulting from a long duration of fading, FEC is useless. The error signals are replaced with the previous one, using an error flag given by the cyclic redundancy check (CRC) error detection. If the error burst becomes very large, we have no option but to blank the signal. Departing from the topic of voice coding, voice activity detection and comfort noise insertion are issues in the voice-operated transmission (VOX) system, where radio signals are transmitted only when the voice signal exists and are turned off during silent periods. Voice activity detection includes methods for measuring the input signal level and level crossing rate, measuring LPC predicted error signal power [107], and for use of the prediction coefficients [105]. With the VOX system, the transmit sound is completely blanked for silent periods. The listener hears no sound, which causes the listener concern that the channel might be disconnected. To remove this, “comfort” noise is generated during silent periods at the receiver [108].

## APPENDIX 7.A AVERAGE ERROR RATE FOR MAXIMAL RATIO COMBINER WITH COHERENT DETECTOR

Average error rate is given by Equation 7.13a as

$$\langle P_e \rangle = \int_0^\infty \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) \frac{\gamma^{M-1}}{(M-1)! \gamma_0^M} e^{-\gamma/\gamma_0} d\gamma$$

Integrating by part, we have

$$\langle P_e \rangle = \left[ \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) I_M(\gamma) \right]_0^\infty - \int_0^\infty \frac{d}{d\gamma} \left( \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) \right) I_M(\gamma) d\gamma \quad (7.A.1)$$

where

$$I_M(\gamma) = \int_0^\infty \frac{\gamma^{M-1}}{(M-1)! \gamma_0^M} e^{-\gamma/\gamma_0} d\gamma = - \sum_{m=0}^{M-1} \frac{\gamma^m}{m! \gamma_0^m} e^{-\gamma/\gamma_0} \quad (7.A.2)$$

We can find

$$\left[ \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) I_M(\gamma) \right]_0^\infty = \frac{1}{2} \quad (M \geq 1) \quad (7.A.3)$$

Using

$$\frac{d}{d\gamma} \left( \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) \right) = -\frac{1}{2} \sqrt{\frac{\alpha}{\pi}} \frac{1}{\sqrt{\gamma}} e^{-\alpha\gamma} \quad (7.A.4)$$

the second term on the right-hand side of Equation 7.A.1 is given using

Equations 7.A.3 and 7.A.4 as

$$\int_0^{\infty} \frac{d}{d\gamma} \left( \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) \right) I_M(\gamma) d\gamma = -\frac{1}{2} \sqrt{\frac{\alpha}{\pi}} \sum_{m=0}^{M-1} J_m \quad (7.A.5)$$

where

$$J_m = \int_0^{\infty} \frac{\gamma^m}{m! \gamma_0^m} \frac{1}{\sqrt{\gamma}} e^{-(\alpha+1/\gamma_0)\gamma} d\gamma \quad (7.A.6)$$

Integrating Equation 7.A.6 by part for  $m \geq 1$ ,

$$J_m = \frac{(2m-1)/2m}{\beta\gamma_0} J_{m-1}$$

where  $\beta = \alpha + 1/\gamma_0$ . Using the relation  $\int_0^{\infty} \frac{1}{\sqrt{\gamma}} e^{-\beta\gamma} d\gamma = \sqrt{\frac{\pi}{\beta}}$ , we get  $J_0 = \sqrt{\pi/\beta}$ , and we have

$$J_m = \frac{(2m-1)!!}{(2m)!!} \frac{1}{(\beta\gamma_0)^m} \sqrt{\frac{\pi}{\beta}} \quad (7.A.7)$$

With use of Equations 7.A.3, 7.A.5, and 7.A.7, Equation 7.A.1 yields

$$\begin{aligned} \langle P_e \rangle &= \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\alpha}{\pi}} \left( J_0 + \sum_{m=1}^{M-1} J_m \right) \\ &= \frac{1}{2} - \frac{1}{2} \frac{1}{\sqrt{1+1/(\alpha\gamma_0)}} \left( 1 + \sum_{m=1}^{M-1} \frac{(2m-1)!!}{(2m)!!} \frac{1}{(1+\alpha\gamma_0)^m} \right) \end{aligned}$$

## APPENDIX 7.B AVERAGE ERROR RATE OF MAXIMAL RATIO COMBINING SYSTEM WITH COHERENT DETECTOR WITH USE OF APPROXIMATE PROBABILITY DENSITY FUNCTION

The error rate is given by

$$\langle P_e \rangle = \int_0^{\infty} \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) \frac{1}{(M-1)!} \frac{\gamma^{M-1}}{\gamma_0^M} d\gamma$$

Integrating by part, have

$$\begin{aligned} \langle P_e \rangle &= \left[ \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) \frac{1}{M!} \frac{\gamma^M}{\gamma_0^M} \right]_0^{\infty} - \int_0^{\infty} \frac{d}{d\gamma} \left( \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}) \right) \frac{1}{M!} \frac{\gamma^M}{\gamma_0^M} d\gamma \\ &= \frac{1}{2} \sqrt{\frac{\alpha}{\pi}} \int_0^{\infty} \frac{1}{\sqrt{\gamma}} e^{-\alpha\gamma} \frac{1}{M!} \frac{\gamma^M}{\gamma_0^M} d\gamma \end{aligned}$$

Comparing this with Equations 7.A.6 and 7.A.7, we get

$$\langle P_e \rangle = \frac{1}{2} \frac{(2M-1)!!}{(2M)!!} \frac{1}{(\alpha\gamma_0)^M}$$

## REFERENCES

- [1] Schwartz M, Benett WR, Stein S. *Communication Systems and Techniques*. New York: Mc Graw-Hill; 1966.
- [2] Jakes WC, editor. *Microwave Mobile Communications*. New York: Wiley; 1974.
- [3] Adachi F, Oono K. BER performance of QDPSK with postdetection diversity reception in mobile radio channels. *IEEE Trans Veh Technol* February 1991;40:237–249.
- [4] Okanou K, Ushirokawa A, Tomita H, Furuya Y. New MLSE receiver free from sample timing and input level controls. *Proceedings of the IEEE Vehicular Technology Conference*; May, 1993. p. 408–411.
- [5] Hattori T, Hirade K. Multitransmitter digital signal transmission by using offset frequency strategy in a land mobile telephone system. *IEEE Trans Veh Technol* November 1978;VT-27:231–238.
- [6] Hattori T, Ogose S. A new modulation scheme for multitransmitter simulcast digital mobile radio communication. *Proceedings of IEEE Vehicular Technology Conference*; March, 1979. p. 83–88.
- [7] Adachi F. Transmitter diversity for a digital FM paging system. *IEEE Trans Veh Technol* November 1979;VT-28:333–338.
- [8] Afrasteh A, Chukurov D. Performance of a novel selection technique in an experimental TDMA system for digital portable radio communications. *Proceedings of IEEE Globecom*; November, 1988. p. 810–814.
- [9] Akaiwa Y. Antenna selection diversity for framed digital signal transmission in mobile radio channel. *Proceedings of IEEE Vehicular Technology Conference*; May, 1989. p. 470–473.
- [10] Barnard JH, Pauw CK. Probability of error for selection diversity as a function of dwell time. *IEEE Trans Commun* August 1989;COM-37:800–803.
- [11] Yamao Y, Nagao Y. Predictive antenna selection diversity (PASD) for TDMA mobile radio. *Trans IEICE* May 1994;E77-B:641–646.
- [12] T. Ohgane and Y. Ogawa, *MIMO System Technology*, Tokyo: Ohmsha, 2009 (in Japanese).
- [13] H. Yanai and K. Takeuchi, *Projection Matrix, Generalized Inverse Matrix, Singular Value Decomposition*, Tokyo: University of Tokyo Press, 1983 (in Japanese).
- [14] S. A. Alamamouti, A simple transmit diversity technique for wireless communications, *IEEE J Sel Area Commun*, October 1998;16(8):1451–1458.
- [15] V. Tarokh, N. Seshadri and A. R. Calderbank, Space-time codes for high data rate wireless communication: performance criterion and code construction, *IEEE Trans Inform Theory*, March 1998;44(2):744–765.
- [16] H. Wang and X. G. Xia, Upper bounds of rates of complex orthogonal space-time block codes, *IEEE Trans Inform Theory*, October 2003;49(10):2788–2796.

- [17] Sampei S. *Application of Digital Wireless Technological to Global Wireless Communications*. Upper Saddle River: Prentice Hall; 1997.
- [18] Proakis G. *Digital Communications*. 3rd ed. New York: McGraw-Hill; 1995.
- [19] E. Viterbo and J. Boutros, A universal lattice code decoder for fading channels, *IEEE Trans Inform Theory*, July 1999;45(5):1639–1642.
- [20] Kim KJ, Yue J. Joint channel estimation and data detection algorithms for MIMO-OFDM systems. *Proceedings of the 36th Asilomar Conference on Signals, Systems and Computers*, Vol. 2; Nov. 2002. p. 1857–1861.
- [21] Kobayashi T, Yano T, Masui H. Computational complexity reduction of maximum likelihood detection for MIMO systems. *Hitachi-Kokusai Technical Report*; No. 8, 2007. p. 58–63.
- [22] M. H. M. Costa, Writing on dirty paper, *IEEE Trans Inform Theory* May 1983;IT-29(3): 439–441.
- [23] Harashima H, Miyakawa H. Matched-transmission technique for channels with Intersymbol Interference. *IEEE Trans Commun*, 1972;COM-20:774–780.
- [24] M. Tomlinson, New automatic equalizer employing modulo arithmetic. *Electron Lett*, March 1971;7(5/6):138–139.
- [25] C. Windpassinger, R. F. H. Fisher, T. Vencel and J. B. Huber, Precoding in multiantenna and multiuser communications, *IEEE Trans Wireless Commun*, July 2004;3(4): 1305–1319.
- [26] Proakis JG. *Digital Communications*. 3rd ed. New York: McGraw-Hill; 1983.
- [27] Sato Y. *Theory of Linear Equalization, Adaptive Digital Signal Processing*. Maruzen: 1990 [in Japanese].
- [28] Lucky RW. Automatic equalization for digital communication. *BSTJ* April 1965;44: 547–588.
- [29] Sato Y. Blind equalization and blind sequence estimation. *IEICE Trans Commun*, May 1994;E77-B:545–556.
- [30] Widrow B, Stearns SD. *Adaptive Signal Processing*. Englewood Cliffs (NJ): Prentice-Hall; 1985.
- [31] Haykin S. *Introduction to Adaptive Filters*. New York: Macmillan; 1984.
- [32] S. Kasturia, J. Aslanis, and J. Cioffi, “Vector coding for partial response channels”, *IEEE Trans Inform Theory*, July 1990;36(4):741–762.
- [33] Falconer D, Ariyavisitakul SL, Benyamin-Seeyar A, Eidson B. Frequency domain equalization for single-carrier broadband wireless systems. *IEEE Commun Mag*, April 2002; 40:58–66.
- [34] C. Douillard, A Picart, P Didier, M Jezequel, C Berrou, A Glavieux, Iterative correction of intersymbol interference: turbo-equalization. *Eur Trans Telecommun Relate Technol*, September–October 1995;6(5):507–511.
- [35] M. Tüchler, R. Koetter, and A.C. Singer, Turbo equalization: principles and new results. *IEEE Trans Commun*, May 2002;50(5):754–767.
- [36] Koetter R, Singer A, Tüchler M. Turbo equalization. *IEEE Signal Process Mag*, January 2004;21:67–80.
- [37] N. Nefedov, M. Pukkila, R. Visoz and A. O. Berthet, Iterative data detection and channel estimation for advanced TDMA systems, *IEEE Trans Commun*, February 2003;51(2): 141–144.

- [38] Namiki J. Block demodulation for short radio packet. IEICE Trans Commun January 1984;J67-B:54–61.
- [39] Hata M, Miki T. Performance of MSK high-speed digital transmission in land mobile radio channels. Proceedings of Globecom '84; November, 1984. p. 518–524.
- [40] Raith K, Stjernvall J-E, Uddenfeldt J. Multi-path equalization for digital cellular radio operating at 300 kbits/s. Proceedings of IEEE Vehicular Technology Conference; 1986. p. 268–272.
- [41] Stjernvall J-E, Hedberg B, Ekemark S. Radio test performance of a narrowband TDMA system. Proceedings of IEEE Vehicular Technology Conference; 1987. p. 293–299.
- [42] Nakajima M, Sampei S. Performance of a decision feedback equalizer under frequency selective fading in land mobile communications. Trans IEICE 1989;72-B-II:515–523.
- [43] Honma K, Uesugi M, Tsubaki K. Adaptive equalization in TDMA digital mobile radio. Trans IEICE 1989;J72-B-II:587–594.
- [44] Ariyavisitakul S Performance bounds for a decision feedback equalizer with a time-reversal structure. Proceedings of the Fourth Nordic Seminar on Digital Mobile Radio Communications; June, 1990. p. 10.a.
- [45] Higashi A, Suzuki H. Dual-mode equalization for digital mobile radio. Transactions IEICE, J74-B-II; March, 1991. p. 91–100.
- [46] Maseng T. Digitally modulated (DPM)signals. IEEE Trans Commun September 1985; COM-33:911–918.
- [47] D'Avella R, Moreno L, Sant'Agostino M. Adaptive equalization in TDMA mobile radio systems. Proceedings of the IEEE Vehicular Technology Conference; May, 1987. p. 385–392.
- [48] Moreno L, D'Arella R. Maximum likelihood adaptive techniques in the digital mobile radio environment. Proceedings of the International Conference on Digital Land Mobile Radio Communications; June, 1987. p. 227–236.
- [49] Stjernvall JE, Hedberg B, Raith K, Backstrom T, Dahl RL, Radio test performance of a narrowband TDMA system-DMS90. Proceedings of the International Conference on Digital Land Mobile Radio Communications; June, 1987. p. 310–318.
- [50] D'Avella R, Moreno L, Sant'Agostino M. An adaptive MLSE receiver for TDMA digital mobile radio. IEEE J Sel Area Commun January 1989;7:122–129.
- [51] Okanou K, Nagata Y, Furuya Y. An adaptive MLSE receiver with carrier frequency estimator for TDMA digital mobile radio. Proceedings of the. Fourth Nordic Seminar on Digital Mobile Radio Communications; June, 1990. p. 10.2.
- [52] Larsson G, Gudmundson B, Raith K. Receiver performance for the north American digital cellular system. Proceedings of the IEEE Vehicular Technology Conference; May, 1991. p. 1–6.
- [53] Kohama T, Kondoh H, Akaiwa Y. An adaptive equalizer for frequency-selective mobile radio channels with noncoherent demodulation. Proceedings of the IEEE Vehicular Technology Conference; May, 1990. p. 770–775.
- [54] Wesley Peterson W, Weldon EJ Jr. *Error-Correcting Codes*. 2nd ed. Cambridge (MA): MIT Press; 1972.
- [55] Lin S, Costello DJ Jr. *Error Control Coding*. Englewood Cliffs (NJ): Prentice-Hall; 1983.
- [56] Lathi BP. *Modern Digital and Analog Communication Systems*. New York: Holt, Rinehart and Winston; 1983.
- [57] Berrou C, Glavieux A, Thitimajhima P. Near shannon limit error-correcting coding and decoding: turbo-codes. Proceedings of the ICC'93; May, 1993. p. 1064–1070.

- [58] C. Berrou, A. Glavieux, Near optimum error correcting coding and decoding: turbo-codes, *IEEE Trans Commun* October 1996;44(10):1261–1271.
- [59] Ogiwara H. *Fundamentals of Turbo Codes*. Tokyo: Triceps; 2008 in Japanese.
- [60] Hanzo L, Liew TH, Yeap BL. *Turbo Coding, Turbo Equalization and Space-Time Coding for transmission over fading channels*. Chichester: IEEE Press/John Wiley & Sons; 2002.
- [61] Bahl LR, Cocke J, Jelinek F, Raviv J. Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Trans Info Theory* 1974;284–287.
- [62] 3GPPTS25.222 (3rd Generation Partnership Project: Technical Specification Group Radio Access Network; Multiplexing and channel coding (TDD) )
- [63] Gallager RG. *Low Density Parity Check Code, in Research Monograph series*. Cambridge: MTT Press; 1963.
- [64] Mackay DJC. Good error-correcting codes based on very sparse matrices. *IEEE Trans Inform Theory* 1999;45:399–431.
- [65] Tanner RM. A recursive approach to low complexity codes. *IEEE Trans Inform Theory* 1981;27:533–547.
- [66] D. Divsalar S. Dolinar and F. Pollara, Iterative turbo decoder analysis based on density evolution, *IEEE J Select Area Commun* May 2001;19(5):891–907.
- [67] H. El Gamal and A. R. Hammons, Analyzing the turbo decoder using the Gaussian approximation, *IEEE Trans Inform Theory*, February 2001;47(2):671–686.
- [68] T. J. Richardson, M. A. Shokrollahi and R. L. Urbanke, Design of capacity-approaching irregular low-density parity-check codes. *IEEE Trans Inform Theory* February 2001; 47(2):619–637.
- [69] Bertsekas D, Gallager R. *Data Networks*. 2nd ed. Englewood Cliffs (NJ): Prentice-Hall; 1992.
- [70] E. Dahlman, S. Parkvall, J. Skold and P. Beming, *3G Evolution HSPA and LTE for Mobile Broadband*. New York: Academic Press, 2007.
- [71] Sato T, Kawabe M, Kato T, Fukusawa A. Composition of robust error control scheme using adaptive coding and its effect on data communication. *Transactions IEICE Japan*, J72-B-I; May, 1989. p. 438–445.
- [72] Ungerboeck G. Trellis-coded modulation with redundant signal sets, Part I: introduction. *IEEE Commun Mag* February 1987;25:5–11.
- [73] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets, Part II: state of the Art", *IEEE Commun Mag*, February 1987; 25, 12–21.
- [74] Murata H, Yoshida S, Takeuchi T. Trellis-coded co-channel interference canceller for mobile communication. Technical Report of IEICE, RCS93-75; November, 1993. p. 39–46.
- [75] Divsalar D, Simon MK. The design of trellis coded MPSK for fading channels: performance criteria. *IEEE Trans Commun* September 1988;36:1004–1012.
- [76] Divsalar D, Simon MK. Set partitioning for optimum code design. *IEEE Trans Commun* September 1988;36:1013–1021.
- [77] Sampei S, Kamio Y. Performance of trellis coded 16QAM/TDMA system for land mobile communications. *Transactions IEICE*, J73-B-II; November, 1990. p. 630–638.
- [78] Tomisato S, Suzuki H. Envelope controlled digital modulation improving power efficiency of transmitter amplification-application to trellis-coded 8PSK for mobile radio. *Transactions IEICE*, J75-B-II; December, 1992. p. 918–928.
- [79] Compton RT Jr. *Adaptive Antennas-Concept and Performance*. Englewood Cliffs: Prentice-Hall; 1988.

- [80] Winter JH. Optimum combining in digital mobile radio with cochannel interference. *IEEE J Sel Area Commun* July 1984;SAC-2:538–539.
- [81] Suzuki H. Signal transmission characteristics of diversity reception with least-squares combining—relationship between desired signal combining and interference cancelling. *Transactions IEICE*, J75-B-II; August, 1992. p. 524–534.
- [82] Yoshino H, Fukawa K, Suzuki H. Adaptive interference canceller based upon RLS-MLSE. *Transactions IEICE*, J77-B-II; February, 1994. p. 74–84.
- [83] Fukawa K, Suzuki H. Recursive least squares adaptive algorithm maximum likelihood sequence estimation: RLS-MLSE—an application of maximum likelihood estimation theory to mobile radio. *Transactions IEICE*, J76-B-II; April, 1993. p. 202–214.
- [84] Yoshida S, Ushirokawa A, Yanagi S, Furuya Y. DS/CDMA adaptive interference canceller in mobile radio environments. Technical Report of IEICE, RCS93-76; November, 1993. p. 47–54.
- [85] Hamaguchi K, Sasaoka H. Block coded FH-16QAM/TDMA system. Technical Report of IEICE, RCS93-90; February, 1994. p. 33–39.
- [86] Kohno R, Imai H, Hatori M, Pasupathy S. Combination of an adaptive array antenna and a canceller of interference for direct-sequence spread-spectrum multiple-access system. *IEEE J Sel Area Commun* May 1990;8:675–682.
- [87] Kohno R, Imai H, Hatori M, Pasupathy S. An adaptive canceller of cochannel interference for spread-spectrum multiple-access communication networks in a power line. *IEEE J Sel Area Commun* May 1990;8:691–699.
- [88] Z. Xie, R.T. Short and C. T. Rushforth, A family of suboptimum detectors for coherent multiuser communications", *IEEE J Sel Area Commun*, May 1990;8:683–690.
- [89] Varanasi MK, Aazhang B. Multistage detection in asynchronous code-division multiple-access communications. *IEEE Trans Commun* April 1990;COM-38:509–519.
- [90] Yoon YC, Kohno R, Imai H. A spread-spectrum multiaccess system with cochannel interference cancellation for multipath fading channels. *IEEE J Sel Area Commun* September 1993;11:1067–1075.
- [91] Abdulrahman M, Falconer DD, Sheikh AUH. Equalization for interference cancellation in spread spectrum multiple access systems. *Proceedings of the IEEE Vehicular Technology Conference*; May, 1992. p. 71–74.
- [92] Madhow U, Honig M. Minimum mean squared error interference suppression for direct-sequence spread-spectrum code division multiple-access. *Proceedings of the First International Conference on Universal Personal Communication*; September, 1992. p. 273–277.
- [93] Fukawa K, Suzuki H. A reception scheme utilizing matched filter with interference cancelling—optimum detection for DS-CDMA mobile communication. *Proceedings of the Spring Conference of IEICE*; March, 1994. p. 1.467–1.468.
- [94] Ozawa K, Miyano T. *Efficient Voice Codings for Digital Mobile Communication*. Tokyo: Triceps; 1992, in Japanese.
- [95] Flanagan JL. *Speech Analysis and Perception*. 2nd ed. New York: Springer-Verlag; 1972.
- [96] Jayant NS, Noll P. *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Englewood Cliffs (NJ): Prentice-Hall; 1984.
- [97] Vary P et al. Speech codec for the European mobile radio system. *Proceedings of the ICASSP*; 1988. p. 227–230.



- [98] Gerson IA, Jasiuk MA. Vector sum excited linear prediction (VSELP) speech coding at 8 kbps. IEEE, Proceedings of the ICASSP; 1990. p. 461–464.
- [99] I. A. Gerson, M. A. Jasiuk, M. J. McLaughlin, and E. H. Winter, "Combined speech and channel coding at 11.2 kbps", *Signal Processing V: Theories and Applications*, In: L. Torres, E. Masgrau and M. A Lagnus, editors. 1339–1342, Elsevier Science Publishers B.V., Amsterdam;1990.
- [100] Ohya T, Suda H, Miki T. Pitch synchronous innovation CELP (PSI-CELP)—PDC half-rate speech CODEC. IEICE Technical Report, RCS 93-78; November, 1993. p. 63–70 [in Japanese]. Also Cox RV. Speech coding standards. In: Kleijn WB, Paliwal KK, editors. *Speech Coding and Synthesis*. Amsterdam: Elsevier Science B.V.; 1995. p. 49–78.
- [101] TIA/ETA/IT-641. Enhanced variable bit rate codec: speech service option 3 for wide spread spectrum digital systems, January 1997.
- [102] Naka N, Ohya T, Saigusa M, Hama T. Special article on mobile multimedia signal processing technologies, speech coding technology. NTT DoCoMo Tech J 2(4):23–32.
- [103] Nakamura O, Dobashi A, Seki K, Kubota S, Kato S. Improved ADPCM voice transmission for TDMA-TDD systems. Proceedings of the Vehicular Technology Conference; May, 1993. p. 301–304.
- [104] Kubota S, Dobashi A, Hasumi T, Suzuki M, Kato S. Improved ADPCM voice transmission employing click noise detection scheme for TDMA-TDD systems. Proceedings of the Fourth International Conference Personal, Indoor and Mobile Radio Communications; September, 1993. p. 613–617.
- [105] Matsumoto I, Sasaki S, Horoguchi M, Urabe K. Enhancement of speech coding for digital cordless telephone systems. Proceedings of the Fourth International Conference Personal, Indoor and Mobile Radio Communications; September, 1993. p. 618–621.
- [106] H. Suda and T. Miki, "An error protected 16 kbits/s voice transmission for land mobile radio", IEEE J Sel Area Commun, 1988;SAC-6:346–352.
- [107] GSM Recommendation 06.32, *Voice activity detection*.
- [108] GSM Recommendation 06.12, *Comfort noise aspects for full rate speech traffic channels*.

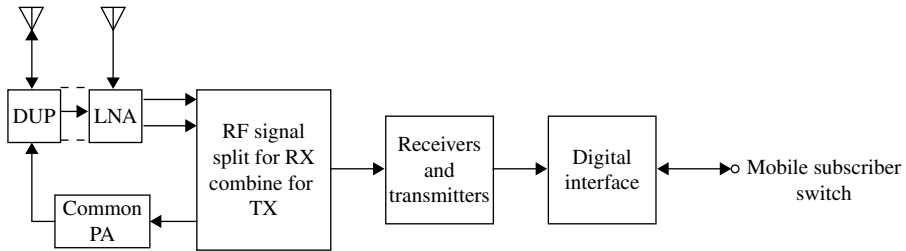
This page intentionally left blank

## EQUIPMENT AND CIRCUITS FOR DIGITAL MOBILE RADIO

This chapter describes equipment and circuit aspects of digital mobile radio systems. The basic structures of the base station and mobile station equipment and some detailed circuit implementations are shown. Some methods and circuits shown here as well as in other parts of this book might be patented. With the evolution of digital signal processing including microcomputers and field programmable gate arrays (FPGAs), many parts of mobile radio systems have been implemented using them. Then the circuit design reduces to programming of them by mathematical expression of operation of the circuit. This topic is not covered in this book. The reader may refer to reference [1].

### 8.1 BASE STATION

Figure 8.1 shows a simplified block diagram of a base station for the Japanese digital cellular system in the NTT Mobile Communications Network Inc. [2]. A base station covers three sectorized zones (Section 9.1.1). Two-branch diversity reception is used. The antenna gains are 17 and 14 dBi at 800 MHz for antenna aperture lengths of 5.4 and 2.7 m, respectively. A transmit and receive duplexer is connected to one antenna, from which signals are transmitted. In order to achieve a low-noise figure, an outdoor common receive low-noise amplifier is used, whose noise figure is less than 3 dB and gain 40 dB. The beam can be tilted by 0–5 or 3–11 degrees in the vertical plane by phase shifting the feeder circuits for the vertical array antenna (Section 9.1.1).



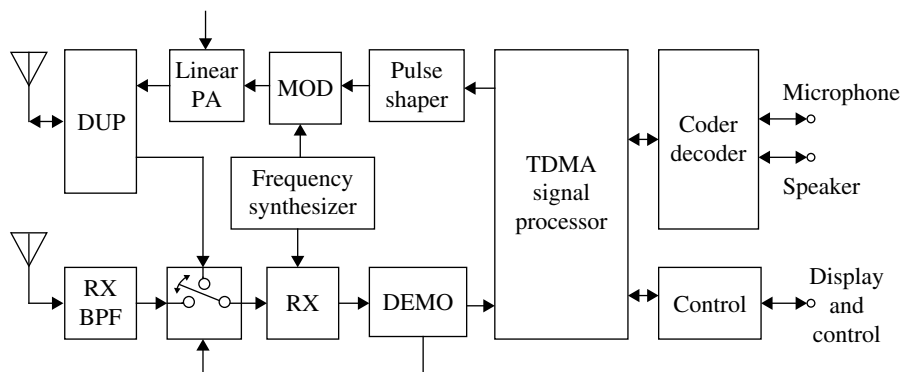
**FIGURE 8.1** Block diagram of base station for a digital cellular system.

The transmit common amplifier can handle a total power of 32 W, accommodating 16 radio-frequency (RF) signals ( $\pi/4$  shifted modulation) and keeping the intermodulation at  $-60$  dB. This performance was achieved by the feed-forward nonlinearity compensation technique, SAFF, described in Section 8.6.3. This common amplifier contributed to a drastic reduction in the size and cost of the base station transmitter: the bulky conventional signal combiner, which is made of many waveguide filters, is replaced with a very small microstrip line combiner. This becomes possible because, due to the common amplifier, the signals are combined at a low level, making them free of transmitter intermodulation. The importance of compactness and low weight of the base station equipment cannot be overstated because of its flexibility for installation at different sites such as roofs of commercial buildings. In addition, channel assignment, such as dynamic channel assignment, becomes flexible since it is free from the restrictions of the conventional channel combining system.

The transmitter/receiver parts of the block diagram deal with modulation/demodulation, TDMA frame composition/decomposition, and so on, with output/input of RF signals and input–output of coded speech signals. The interface circuit covers multiplexed transmission of coded speech signals and other signals for control.

## 8.2 MOBILE STATION

A block diagram of a handheld mobile station for the 800 MHz Japanese digital cellular system is shown in Figure 8.2 [3, 4]. Antenna selection diversity is used (Section 7.1.4). One is a whip antenna used for transmission and reception via a duplexer. The other is an inverse F-type antenna imbedded on the housing and is used for only reception. Antenna selection is controlled by the output of the demodulator. A three-slot TDMA signal is pulse shaped, modulated with  $\pi/4$  shifted QPSK, and amplified to a maximum output power of 0.8 W. The output power is automatically controlled. The envelope of the modulating signal is fed to the power amplifier for nonlinearity compensation purposes. The 11.2 kbps VSELP voice coder/decoder (Section 7.7.6) is implemented with a one-chip digital signal processor. Voice-operated transmission (VOX, Section 7.7.8) and intermittent reception are adopted, decreasing the power consumption. The battery life on each charging is about 30 h in the waiting mode and 60 min in the speaking mode at maximum power without the VOX.



**FIGURE 8.2** Block diagram of a mobile station for a digital cellular system.

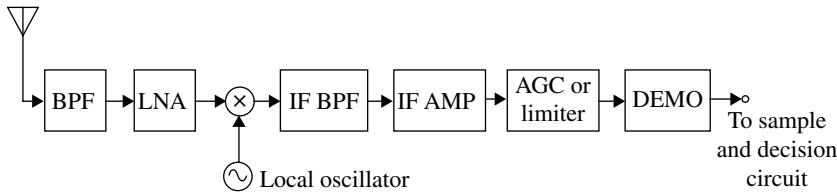
The receive circuit (RX) includes an RF low-noise amplifier, a frequency down-converter, a channel selection band-pass filter, and an intermediate frequency (IF) amplifier. A carrier frequency is generated using a frequency synthesizer. Fast switching of less than 2 ms is achieved with a digital loop reset technique.

The two-receiver postmodulation diversity system is adopted for the car-mounted and the shoulder-type mobile stations.

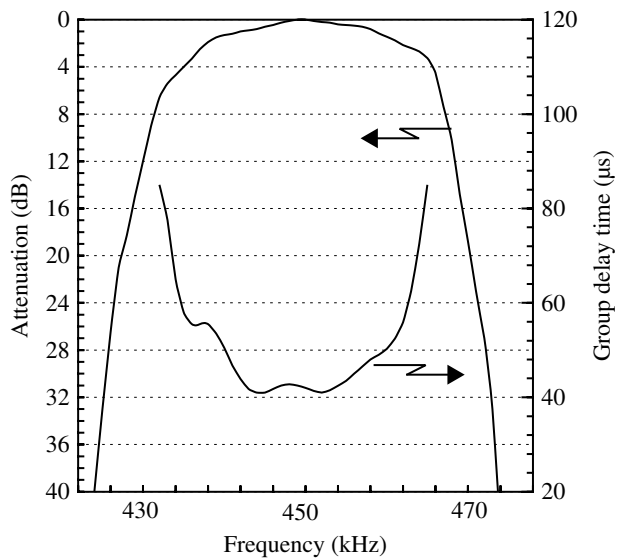
### 8.3 SUPERHETERODYNE AND DIRECT CONVERSION RECEIVERS

A block diagram of the superheterodyne receiver is shown in Figure 8.3. The RF received signal is amplified and frequency downconverted into an IF signal. The gain of the low-noise amplifier should not be high so as to prevent intermodulation between the desired signal and others. The band-pass filter is used for channel selection. A steep roll-off transfer function is desired for the channel selection capability. An example of the square root of the Nyquist I transfer function of the channel selection filter for the American TDMA digital cellular system is shown in Figure 8.4. The delay and amplitude characteristics are taken into consideration for low distortion of the modulated signal waveform. A high gain is implemented for the IF amplifier since a channel is already selected, and therefore no intermodulation is expected. The output of the IF amplifier is fed to an amplitude limiter or an automatic gain control (AGC) circuit followed by a demodulator. A further frequency conversion is sometimes made (double-conversion system).

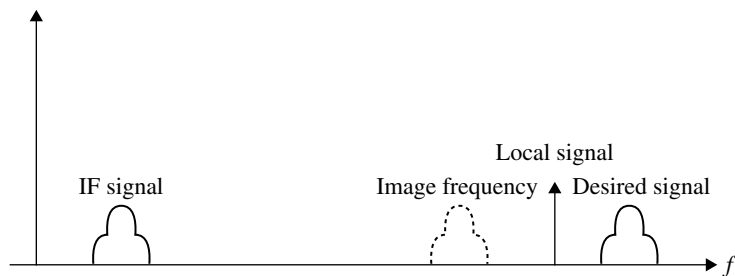
The advantages of the heterodyne receiver are high channel selectivity owing to moderate signal bandwidth relative to the IF, and stable, high-gain amplification owing to the frequency difference between the input RF signal and the IF signal. The disadvantage of the heterodyne receiver is the image frequency problem: undesired signal that is located at a frequency in mirror-image relation to the desired signal with



**FIGURE 8.3** Block diagram of a superheterodyne receiver.

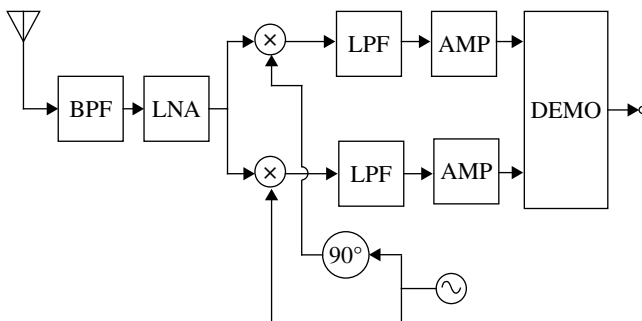


**FIGURE 8.4** Frequency characteristics of a ceramic band-pass filter for the US digital cellular system. Courtesy of Murata Manufacturing Co. Ltd.



**FIGURE 8.5** Image frequency for a heterodyne receiver.

respect to the local frequency (Fig. 8.5) may be downconverted to the same IF and not discriminated with the channel selection filter. This is known as the image frequency problem. To avoid this problem, a band-pass filter that attenuates the image frequency signal is employed before the frequency downconversion.



**FIGURE 8.6** Block diagram of a direct conversion receiver.

The direct conversion, or homodyne, receiver block diagram is shown in Figure 8.6. The RF signal is downconverted into quadrature baseband signals using a pair of downconverters with a local signal whose frequency is the same as the carrier signal. The channel selection is made with the low-pass filters (LPFs). Since the LPFs and the following circuits treat baseband signal, they are suited for implementation with integrated circuits. Moreover, we have no image frequency problem. However, we have new problems: (i) the  $1/f$  noise appears in baseband frequency, (ii) dc-coupled circuits are required for baseband signals, and (iii) leaked local signals easily interfere with other receivers located close-by.

The principle of demodulation for the direct conversion signal is based on the zero IF complex or quadrature representation of modulated signals. Consider an FM signal, for example; the RF signal is represented as

$$\begin{aligned} s(t) &= A_0 \cos[\omega_c t + \theta(t)] \\ &= \text{Re} \left[ A_0 e^{j\theta(t)} e^{j\omega_c t} \right] \end{aligned}$$

where  $\omega_c$  is the carrier frequency and  $\theta(t) = k \sum_{-\infty}^t f(t) dt$ , where  $f(t)$  is the transmit baseband signal and  $k$  is the modulation constant. The demodulated quadrature signals become  $x(t) = A_0 \cos \theta(t)$  and  $y(t) = A_0 \sin \theta(t)$ . The demodulator works on  $x(t)$  and  $y(t)$  as

$$\begin{aligned} d(t) &= \frac{d}{dt} \theta(t) = \frac{d}{dt} \tan^{-1} \frac{y(t)}{x(t)} = \left[ x(t) \frac{dy(t)}{dt} - y(t) \frac{dx(t)}{dt} \right] / A_0^2 \\ &= kf(t) \end{aligned}$$

### 8.3.1 Image Rejection Downconverter

To cope with the difficulties with direct conversion receivers, heterodyne receivers with a low IF is known. However, the image frequency problem still exists for the receiver, since they are heterodyne receivers. Here, we discuss low IF receivers that suppress the image frequency component with a quadrature demodulator.

The principle of this system is based on single sideband conversion (Section 2.3.4, Example 2.10).

Any modulated signal is expressed as

$$s(t) = x(t) \cos \omega_c t - y(t) \sin \omega_c t \quad (8.1)$$

where  $x(t)$ ,  $y(t)$ , and  $\omega_c$  are inphase, quadrature components, and carrier frequency, respectively. We have complex expression as

$$s(t) = \text{Re} \left\{ z(t) e^{j\omega_c t} \right\}$$

where

$$z(t) = x(t) + jy(t)$$

Quadrature detection of the signal with Equation 8.1 yields output signal except for constant as

$$\begin{aligned} a(t) &= x(t) \cos \omega_{\text{IF}} t - y(t) \sin \omega_{\text{IF}} t \\ b(t) &= -x(t) \sin \omega_{\text{IF}} t - y(t) \cos \omega_{\text{IF}} t \end{aligned}$$

where  $\omega_{\text{IF}}$  is the IF. That is,  $\omega_{\text{IF}} = \omega_c - \omega_L$ .  $a(t)$  and  $b(t)$  are expressed as

$$\begin{aligned} a(t) &= \text{Re} \left\{ z(t) e^{j\omega_{\text{IF}} t} \right\} \\ b(t) &= \text{Re} \left\{ z(t) e^{j(\omega_{\text{IF}} t + \pi/2)} \right\} \end{aligned}$$

If  $\omega_{\text{IF}} = -(\omega_c - \omega_L) = \omega_L - \omega_c$ , we have

$$\begin{aligned} a(t) &= x(t) \cos \omega_{\text{IF}} t + y(t) \sin \omega_{\text{IF}} t \\ &= \text{Re} \left\{ z(t) e^{-j\omega_{\text{IF}} t} \right\} = \text{Re} \left\{ z^*(t) e^{j\omega_{\text{IF}} t} \right\} \\ b(t) &= x(t) \sin \omega_{\text{IF}} t - y(t) \cos \omega_{\text{IF}} t \\ &= \text{Re} \left\{ z(t) e^{-j(\omega_{\text{IF}} t - \pi/2)} \right\} = \text{Re} \left\{ z^*(t) e^{j(\omega_{\text{IF}} t - \pi/2)} \right\} \\ &= -\text{Re} \left\{ z^*(t) e^{j(\omega_{\text{IF}} t + \pi/2)} \right\} \end{aligned}$$

As input to a quadrature demodulator, we assume two signals  $s(t)$  and  $i(t)$  as

$$\begin{aligned} s(t) &= x_s(t) \cos \omega_s t - y_s(t) \sin \omega_s t \\ i(t) &= x_i(t) \cos \omega_i t - y_i(t) \sin \omega_i t \end{aligned}$$

We also assume that  $\omega_s$  and  $\omega_i$  are in relation to image frequency:

$$\omega_{\text{IF}} = \omega_s - \omega_L = \omega_L - \omega_i$$



Then following the aforementioned arguments, we have

$$\begin{aligned} a(t) &= x_s(t) \cos \omega_{\text{IF}} t - y_s(t) \sin \omega_{\text{IF}} t + x_i(t) \cos \omega_{\text{IF}} t + y_i(t) \sin \omega_{\text{IF}} t \\ &= \text{Re} \left\{ \left[ z_s(t) + z_i^*(t) \right] e^{j\omega_{\text{IF}} t} \right\} \end{aligned} \quad (8.2)$$

$$\begin{aligned} b(t) &= -x_s(t) \sin \omega_{\text{IF}} t - y_s(t) \cos \omega_{\text{IF}} t + x_i(t) \sin \omega_{\text{IF}} t - y_i(t) \cos \omega_{\text{IF}} t \\ &= \text{Re} \left\{ \left[ z_s(t) - z_i^*(t) \right] e^{j(\omega_{\text{IF}} t + \pi/2)} \right\} \end{aligned} \quad (8.3)$$

where we put  $z_s(t) = x_s(t) + jy_s(t)$  and  $z_i(t) = x_i(t) + jy_i(t)$ .

**8.3.1.1 Phase Shift Method** The block diagram of this system is shown in Figure 8.7.

By phase shifting of signal  $b(t)$  (Eq. 8.3) by  $-90^\circ$ , we have

$$\hat{b}(t) = \text{Re} \left\{ \left[ z_s(t) - z_i^*(t) \right] e^{j\omega_{\text{IF}} t} \right\}$$

Comparing  $a(t)$  with  $\hat{b}(t)$ , we see that the desired and the image signal are separately obtained in Figure 8.7. The  $-90^\circ$  phase shift is called the Hilbert transform (Section 2.3.4).

**8.3.1.2 Double Conversion Method** This system is shown in Figure 8.8. If we denote output signals at the second downconversion by  $a'(t)$  and  $b'(t)$ , following the aforementioned arguments, we have

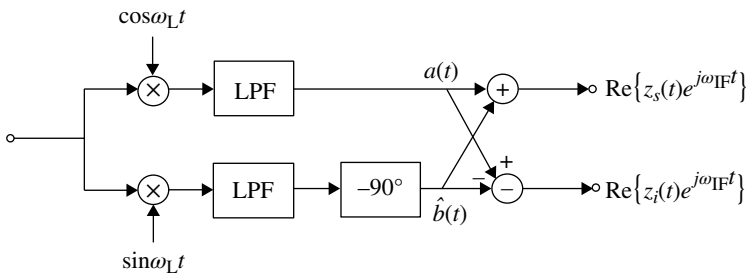
$$a'(t) = \text{Re} \left\{ \left[ z_s(t) + z_i^*(t) \right] e^{j\omega_{\text{IF}} t} \right\} \quad (8.4a)$$

$$b'(t) = -\text{Re} \left\{ \left[ z_s(t) - z_i^*(t) \right] e^{j\omega_{\text{IF}} t} \right\} \quad (8.4b)$$

where  $\omega'_{\text{IF}} = \omega_{\text{IF}} - \omega'_L$ .

We understand from Equations 8.4a and 8.4b that the desired and image signals are separated in Figure 8.8.

Now, we understand the second quadrature mixer is equivalent to the  $-90^\circ$  phase shifting circuit in Figure 8.7 (refer to Example 2.10).



**FIGURE 8.7** Phase shift method.

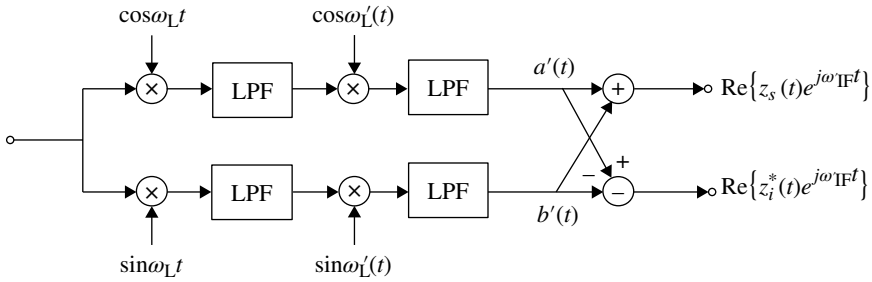


FIGURE 8.8 Double-conversion method.

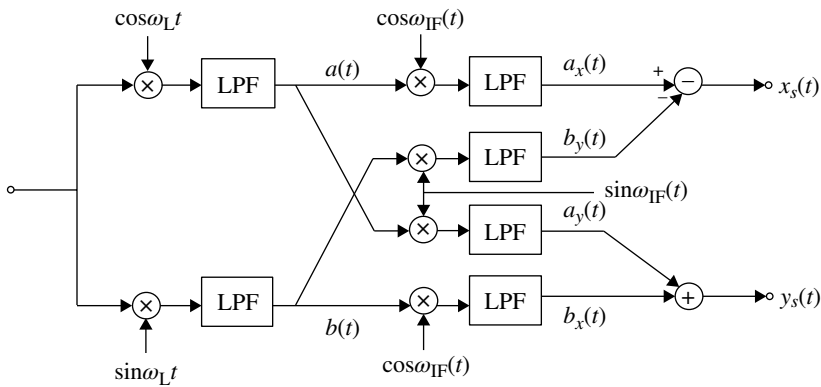


FIGURE 8.9 Modification of the double-conversion method.

**Modification of the Double-Conversion Method** Performing direct conversion of  $a(t)$  and  $b(t)$  to get the inphase and the quadrature components, we have a system shown in Figure 8.9. The last four LPFs may be replaced by two LPFs at the outputs of the addition and subtraction circuit.

From Equations 8.2 and 8.3, signals  $a_x(t)$ ,  $a_y(t)$ ,  $b_x(t)$ , and  $b_y(t)$  in the figure are expressed as

$$2a_x(t) = x_s(t) + x_i(t)$$

$$2a_y(t) = -y_s(t) + y_i(t)$$

$$2b_x(t) = -y_s(t) - y_i(t)$$

$$2b_y(t) = -x_s(t) + x_i(t)$$

From these equations, we get

$$a_x(t) - b_y(t) = x_s(t)$$

$$a_x(t) + b_y(t) = x_i(t)$$

$$a_y(t) + b_x(t) = -y_s(t)$$

$$a_y(t) - b_x(t) = y_i(t)$$

## 8.4 TRANSMIT AND RECEIVE DUPLEXING

An antenna is used for signal transmission and reception via a duplexer. Figure 8.10 shows a block diagram of a duplexer. The center frequencies of the band-pass filters are chosen at the transmit and receive frequency bands, respectively. The circulator is a nonreciprocal circuit where the input signal is fed out to the next port in a rotational manner. The circulator is sometimes removed, connecting the band-pass filters in parallel to the antenna.

The same frequency is sometimes used for transmission and reception, that is, a half-duplex (press-to-talk) system and a time division duplex (TDD) system. Transmission and reception are controlled by pressing or releasing a control button in the press-to-talk system. Time is divided in turn for transmission and reception in the TDD or “ping-pong” system. In these systems, the duplexer is replaced with a switch, reducing the size and cost of the transmitter and receiver, since signal transmission and reception never take place at the same time.

## 8.5 FREQUENCY SYNTHESIZER

For a system in which a number of frequency channels can be selected, a frequency synthesizer [5] is indispensable. A block diagram of a frequency synthesizer is shown in Figure 8.11. This is a phase-locked loop (PLL) circuit, where the voltage-controlled

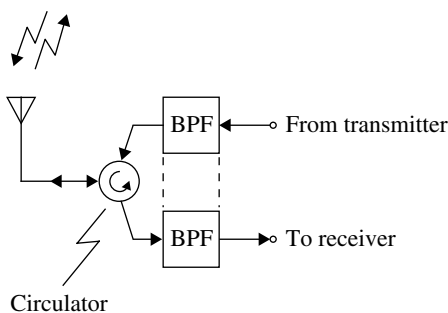


FIGURE 8.10 Antenna and duplexer.

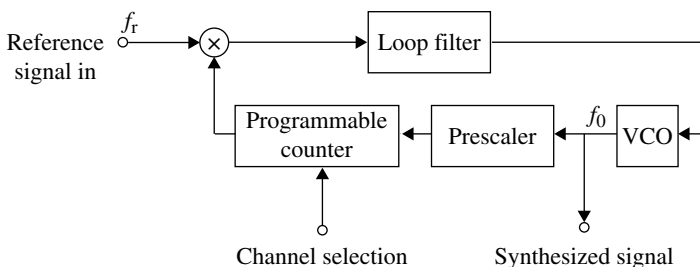


FIGURE 8.11 Frequency synthesizer.

oscillator (VCO) output signal is frequency divided and phase locked to a stable reference signal. The VCO output signal frequency  $f_0$  is determined as

$$f_0 = Nf_r$$

where  $N$  is the total number of frequency divisions and  $f_r$  is the reference signal frequency. If  $f_r = 12.5$  kHz,  $f_0$  can be controlled by a step of 12.5 kHz, for example, from 800 to 825 MHz by letting  $N = 64,000$  to  $N = 66,000$ . The reference signal is generated by a stable quartz oscillator followed by a counter. Using temperature-dependent compensation technique, frequency stability of 1.5 ppm (part per million) is achieved. Frequency division is implemented with the so-called prescaler and the programmable counter.

The former is a fast counter with a fixed counting-up number, and the latter is a frequency-dividing programmable counter where the counting-up number is controlled from the channel selection digital signal. The prescaler is used to cope with high frequencies at low power consumption. A GaAs circuit is suitable for this device.

Compactness, low power consumption, low noise, frequency stability, and fast channel switching are major technical issues for frequency synthesizers in mobile radio equipment. The compactness is governed by the VCO resonator and other electronic circuits. A high-Q resonator and a low-noise transistor are required for low-noise performance. Improvements in transistor technology, large-scale integrated circuits, and high-dielectric substrates for the resonator yield compactness and low-noise performance.

The prescaler is the main power-consuming circuit, since it operates at the highest frequency; the other circuits work at low frequencies, where low-power-consuming complementary metal-oxide-semiconductor (CMOS) transistors can be used. With the improvements in semiconductor technology, the power consumption of the 800 MHz prescaler is reduced to 10 mW for GaAs circuits and 20 mW for more economical Si-bipolar transistor circuits. Intermittent operation of a synthesizer lowers the power consumption. To make this technique effective, the rise-up time must be short. For this purpose, a method termed state-preserving intermittently locked loop (SPILL) is proposed, where the states of the programmable counter and the prescaler are stored just before intermittent power-down periods and are used for the initial states at the time of restarting operation [6].

The stability of the frequency synthesizer is determined by the stability of the reference signal. A high-frequency stability is achieved at the base station, since the requirements for size, cost, and power consumption are less stringent than those at the mobile station. One technique is to adjust the frequency of the reference signal with reference to the received signal frequency transmitted from the base station. Using this kind of technique, a stability of less than 0.3 ppm is achieved.

Fast switching from one frequency to another is required at the moment of hand-off in a cellular system. This is more important for a TDMA system since a mobile station must switch its frequencies to monitor other channels (mobile-assisted hand-off; Section 9.1.1) in short time slots and come back again to the dedicated frequency and time slot. For this purpose, two synthesizers are prepared in a primitive method. A single synthesizer with a fast switching time of less than 1 ms is proposed. In this

method, fast and slow counters are employed as phase detectors (comparators), and the expected dc voltage is added to the control voltage of the VCO. The fast and slow counters cover fine and coarse phase error detection, respectively.

## 8.6 TRANSMITTER CIRCUITS

### 8.6.1 Digital Signal Waveform Generator

In general, digital signal waveforms are generated by applying impulses to a transmit filter. In addition to this, a table lookup method can be used as shown in Figure 8.12. The waveforms are calculated for the data sequences in advance and stored in read-only memory (ROM) and read out corresponding to the address (data sequence) entered into the shift register. The contents of memory are

$$s(n\Delta T) = \sum_{m=-M}^M a_m h(n\Delta T - mT) w(n\Delta T - mT), \quad n = 0, 1, 2, \dots, N-1$$

where  $a_m$  are digital data;  $T$  is the symbol duration;  $\Delta T$  is the sampling interval;  $N = T/\Delta T$ ;  $h(t)$  is the impulse response, which is truncated within  $(2M+1)T$ ; and  $w(t)$  is the window function (Section 2.4.9). For a two-level polar signal,  $a_n$  is +1 or -1. For example, for a square root of the Nyquist I pulse waveform with roll-off factor  $\alpha$  and symbol frequency  $f_s$ , we have

$$h(t) = \frac{f_s}{\pi f_s t} \sin[(1-\alpha)\pi f_s t] + \frac{f_s}{(\pi/4\alpha)^2 - (\pi f_s t)^2} \times \left\{ \frac{\pi}{4\alpha} \cos[(1+\alpha)\pi f_s t] + \pi f_s t \sin[(1-\alpha)\pi f_s t] \right\} \quad (8.5)$$

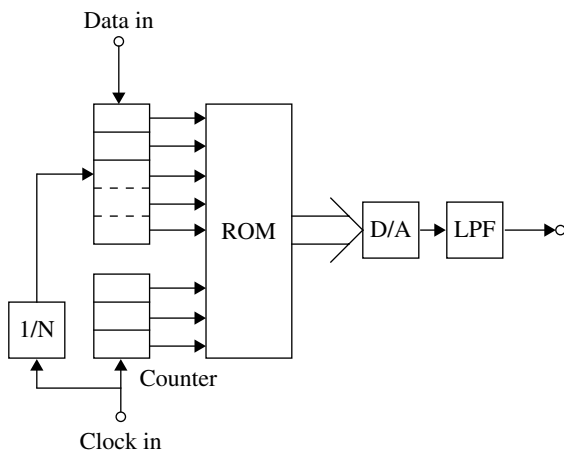


FIGURE 8.12 Digital signal waveform generator with table lookup.

For a rectangular window,

$$w(t) = \begin{cases} 1 & (-MT \leq t \leq MT) \\ 0 & (\text{otherwise}) \end{cases}$$

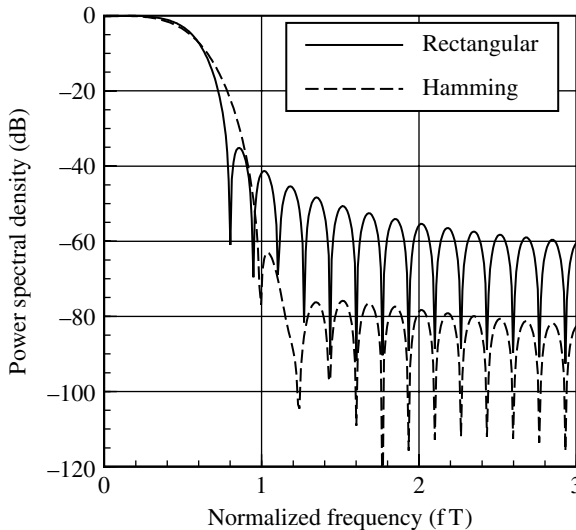
For the generalized Hamming window,

$$w(t) = \begin{cases} \beta + (1 - \beta) \cos\left(\frac{\pi t}{2MT}\right) & (-MT \leq t \leq MT, \beta = 0 \sim 1) \\ 0 & (\text{otherwise}) \end{cases}$$

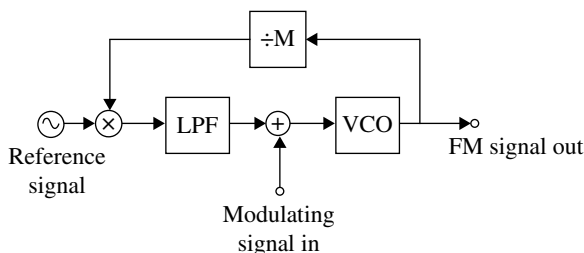
In the aforementioned system, the required number of addresses for ROM becomes  $N 2^{2M+1}$ . The effect of truncating the impulse response and the window function ( $\beta=0.56$ ) is shown in Figure 8.13. The out-of-band radiation is decreased by the introduction of the window function.

## 8.6.2 Modulator

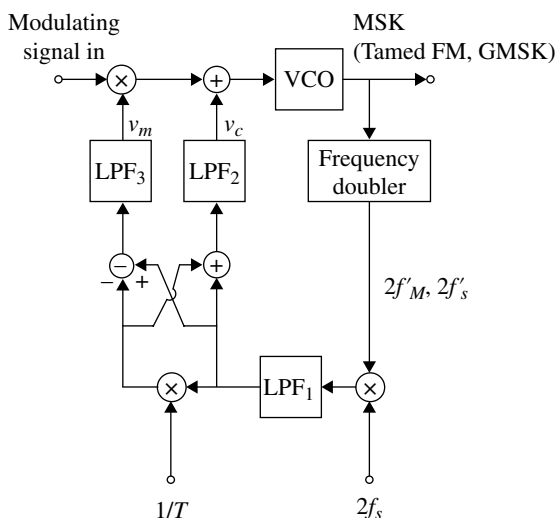
**8.6.2.1 FSK Modulator** A VCO is a typical frequency modulator. For a frequency modulator, stability of center frequency and modulation index is important. Figure 8.14 shows a frequency modulator with center frequency stabilization. The center frequency is locked to a reference signal. To reduce the effect of frequency modulation on the center frequency stabilization, the VCO output signal is frequency divided, resulting in a very small modulation index. This circuit is appropriate for



**FIGURE 8.13** The effect of truncation and windowing on spectra ( $M=3$ ).



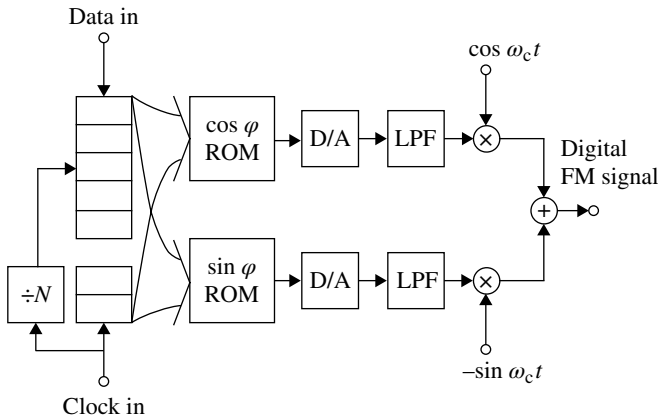
**FIGURE 8.14** Center frequency stabilized FM modulator.



**FIGURE 8.15** Center frequency and modulation index stabilized FM modulator for FSK signals with modulation index of 0.5.

modulating signals without dc components such as analog voice signals and Manchester coded signals (Section 3.2.4).

Figure 8.15 shows an FSK modulator [7] with stabilization of both center frequency and modulation index using the properties of the family of FSK signals (i.e., MSK, tamed FM, and GMSK): the modulation index is 0.5 for these signals. As discussed in Section 6.2.1, discrete frequency components of  $2f'_M = 2(f'_c - \Delta f')$  and  $2f'_S = 2(f'_c + \Delta f')$  are obtained by passing the signal into a frequency doubler, where  $f'_c$  is the center frequency and  $\Delta f'$  is the frequency deviation, which should be  $\Delta f' = 1/4T$  ( $1/T = f_b$ : bit rate frequency).  $f'_c$  and  $\Delta f'$  are assumed to be deviated slightly from the desired frequencies  $f_c$  and  $1/4T$ . In Figure 8.15,  $\text{LPF}_1$  rejects the higher-order carrier frequencies and  $\text{LPF}_2$  and  $\text{LPF}_3$  pass the dc component. When the loops are locked, we can see that the center frequency control signal  $v_c$  is produced so that  $f'_c + f_b/2 = f'_s \rightarrow f_s$  and the gain or (modulation index) control signal  $v_m$  is produced so that  $\Delta f' \rightarrow 1/4T$ .



**FIGURE 8.16** Digital FM signal generator with quadrature modulation.

An FM modulator with quadrature modulation is shown in Figure 8.16. The principle of this modulator is described by the mathematical expression

$$\begin{aligned} s(t) &= \cos[\omega_c t + \varphi(t)] \\ &= \cos \varphi(t) \cos \omega_c t - \sin \varphi(t) \sin \omega_c t \end{aligned}$$

where  $\omega_c$  is the carrier frequency and  $\varphi(t)$  is the phase. The phase  $\varphi(t)$  is given by

$$\varphi(t) = k \sum_{m=-M}^M a_m g(t - mT)$$

where  $k$  is the modulation constant and  $g(t)$  is the phase impulse response

$$g(t) = \int_{-\infty}^t h(x) dx$$

and  $h(t)$  is the frequency impulse response.

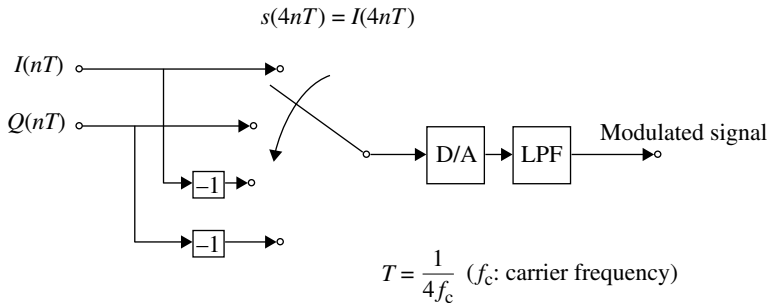
The center frequency and modulation index are kept stable with this circuit. However, balancing the quadrature modulation and spurious carrier frequency component becomes a problem in this system.

The aforementioned problems can be avoided by introducing digital signal processing to the modulator at an IF. However, power-consuming digital multipliers and adders are required if the analog circuits are replaced with a digital signal processor. A simple method is proposed in which signal processing effort is minimum, as shown in Figure 8.17 [8].

A sampled quadrature modulated signal is expressed as

$$s(nT) = I(nT) \cos(\omega_c nT) + Q(nT) \sin(\omega_c nT) \quad n = 0, 1, 2, \dots$$





**FIGURE 8.17** Digital signal processing quadrature modulator.

If we choose  $T = \pi/2\omega_c = 1/4f_c$  ( $f_c = \omega_c/2\pi$ ), we have

$$\begin{aligned} s(4nT) &= I(4nT) \\ s[(4n+1)T] &= Q[(4n+1)T] \\ s[(4n+2)T] &= -I[(4n+2)T] \quad \text{and} \\ s[(4n+3)T] &= -Q[(4n+3)T]. \end{aligned}$$

No multipliers nor adders are used in this system.

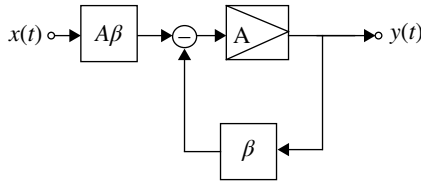
### 8.6.3 Linear Power Amplifier

A power amplifier with high power efficiency and low nonlinear distortion is a key requirement for linear modulation to be feasible in mobile radio communications. Many efforts on nonlinearity compensation methods had been devoted to power-efficient amplifiers. For AM broadcasting and SSB radio communications, nonlinearity compensating techniques have long been investigated [9–12]. Nonlinearity compensated amplifiers can be classified into feedback, predistortion, and feed-forward amplifiers. Here, we explain these techniques placing an emphasis on the predistortion amplifiers, which became more important for application to base station transmitter.

**8.6.3.1 Feedback Amplifier** Feedback technique is well known for nonlinear compensation in power amplifiers, especially for audio amplifiers. Its principle is as follows. Let  $x(t)$  denote the input signal and  $y(t)$  the output signal, and then we have

$$y(t) = Ax(t) + \Delta f_a[x(t)] \quad (8.6)$$

where  $A$  is the (linear) amplification coefficient and the function  $\Delta f_a[\bullet]$  denotes the nonlinear distortion.  $\Delta f_a[\bullet]$  depends on the nonlinearity of the amplifier for a given  $x(t)$ . Let us consider a feedback amplifier shown in Figure 8.18, where the coefficient  $A\beta$  is introduced to keep the output power level unchanged with the feedback control. We have



**FIGURE 8.18** Negative feedback amplifier.

$$y(t) = A[A\beta x(t) - \beta y(t)] + \Delta f_d [A\beta x(t) - \beta y(t)] \quad (8.7)$$

or

$$y(t) = \frac{A^2 \beta x(t) + \Delta f_d [\varepsilon(t)]}{1 + A\beta} \quad (8.8)$$

where  $\varepsilon(t) = A\beta x(t) - \beta y(t)$ .

For  $|A\beta| \gg 1$ ,  $y(t)$  is approximated as

$$y(t) \approx Ax(t) + \frac{\Delta f_d [\varepsilon(t)]}{A\beta} \quad (8.9)$$

With low distortion, that is,  $\Delta f_d [\varepsilon(t)] \ll |A^2 \beta x(t)|$ , using Equation 8.8,  $y(t)$  is approximated as

$$y(t) \approx \frac{A^2 \beta x(t)}{1 + A\beta}$$

and  $\varepsilon(t)$  is approximated as

$$\varepsilon(t) \approx x(t) \quad (8.10)$$

This approximation means that the nonlinear distortion component is neglected in the feedback path. Noting Equation 8.10 and comparing Equations 8.6 and 8.9, we can see that the nonlinear distortion is decreased by a factor of  $A\beta$ , the feedback gain factor.

Feedback control at RF frequencies becomes difficult because of the feedback circuit implementation. To overcome the difficulty, the RF signal is demodulated to produce baseband signals. The baseband signals are negatively fed back to the baseband signals or to modulating circuits. The demodulation of an RF signal in Cartesian coordinates (inphase and quadrature components) and polar coordinate (amplitude and phase components) is shown in Figures 8.19 and 8.20, respectively. The Cartesian loop (feedback control) method is suitable for the quadrature modulator since quadrature components are used. This method was applied to the 140 MHz class AB power amplifier with  $\pi/4$  shifted QPSK [13]. A power efficiency of 35% and out-of-band radiation density of  $-60$  dB are obtained with a feedback gain of 30 dB. The results with this method applied to an 800 MHz power amplifier are reported in Ref. [14].

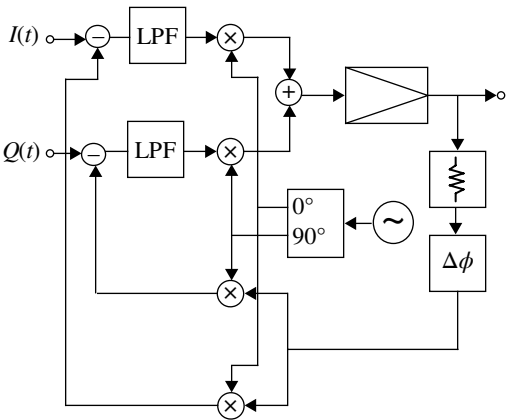


FIGURE 8.19 Cartesian loop modulation feedback amplifier.

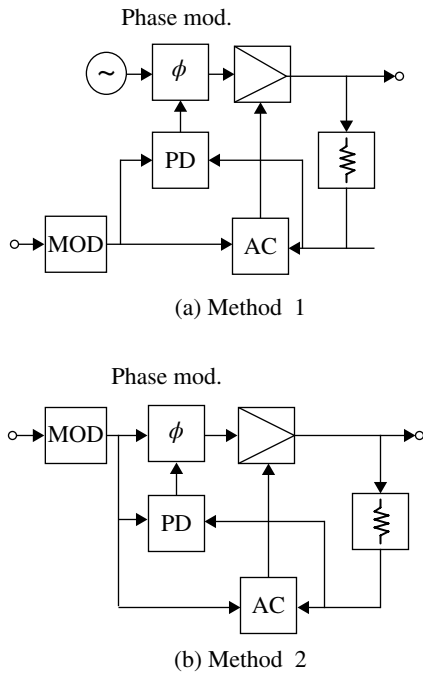


FIGURE 8.20 Polar coordinate demodulation feedback amplifier. PD, phase detector; AC, amplitude comparator.

The polar loop method was applied to a  $\pi/4$  shifted QPSK transmitter [15]. This method has advantages since it is immune to the signal phase shift caused while selecting different channel frequencies in multichannel systems. Results when only the amplitude component is fed back are reported in Ref. [16]. An 850MHz power amplifier with power efficiency of 50% and third-order distortion of  $-30$  dB is obtained.

The advantage of the feedback method is that nonlinearity compensation performance can be maintained even when the amplifier parameters deviate due to temperature and age. The inherent disadvantage for the feedback control is the problem of possible instability (Section 2.3.2). In this application, care must be taken especially, when the channel frequencies are widely varied for channel selection.

**8.6.3.2 Feed-Forward Method** The block diagram of a feed-forward amplifier is shown in Figure 8.21. The distortion or error signal produced in the amplifier is detected by comparing the input and output signals. The detected error signal is fed into the linear subamplifier to amplify it to the same level as that of the main power amplifier. The amplified error signal is subtracted from the output of the power amplifier. The advantages and disadvantages of the feed-forward amplifier are the same as those of the predistorted amplifier.

The adjustment of the delay time and gain in the second path is important for the success of the feed-forward amplifier. The linearity of the subamplifier must be high. This can decrease the overall power efficiency. An automatic error-correcting technique for the feed-forward amplifier was introduced as shown in Figure 8.22.

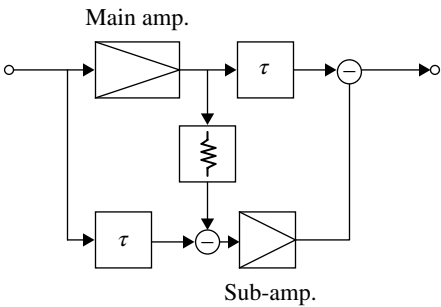


FIGURE 8.21 Feed-forward amplifier.

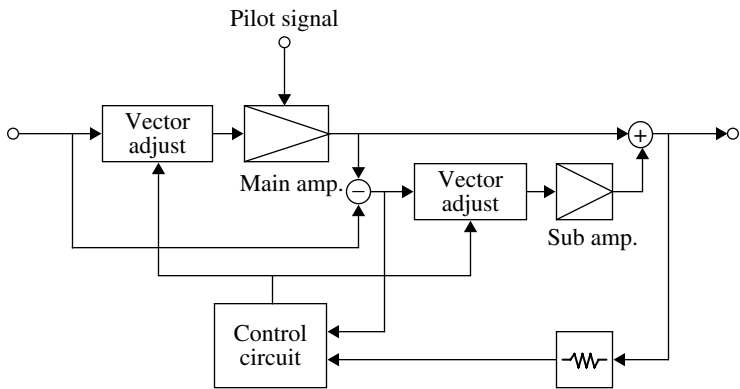
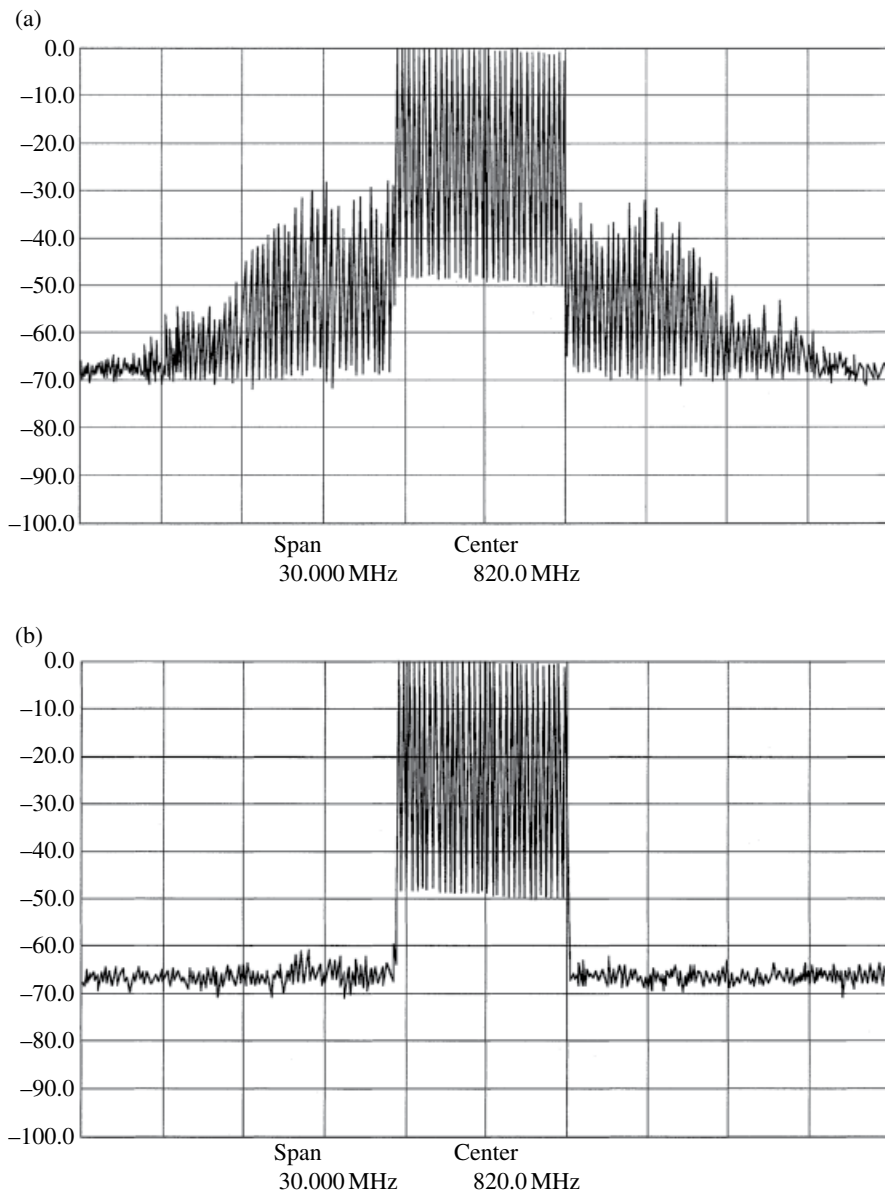


FIGURE 8.22 Adaptive feed-forward amplifier.

Figure 8.23 shows results for an 800 MHz power amplifier [17]. The nonlinear distortion is suppressed more than 30 dB with a 32 W (average) FET power amplifier. This result was obtained over a 20 MHz bandwidth. With this amplifier, experiments were performed on simultaneous amplification of 32 carriers, where each carrier has



**FIGURE 8.23** Spectrum improvement with feed-forward nonlinearity compensation [17]. (a) Without equalizer and (b) with equalizer. Reproduced by permission of Copyright © 1992, NTT.

1 W of power and is spaced at 100 kHz intervals. Leakage power into unused channels was less than  $-60$  dB. These results show that this amplifier is suitable for a multicarrier amplifier that has application in cellular base stations.

In Ref. [10], linearly modulated signals are split into the amplitude component and the phase component with constant envelope and those are amplified independently. The constant envelope phase component is modulated by the amplitude components. This method falls into the category of feed-forward amplifiers. The gain control is actually performed by controlling the supplied voltage of the power amplifier. Using a similar idea, another method controls the bias voltage corresponding to the input signal level.

**8.6.3.3 Linear Amplification with Nonlinear Components Amplifier** The linear amplification with nonlinear component (LINC) system [18] is a combination of the principles of the predistorter and feed-forward techniques (Fig. 8.24). The input signal is generally given by

$$s_a(t) = E(t) \cos[\omega_0 t + \theta(t)]$$

where the envelope  $E(t)$  and phase  $\theta(t)$  correspond to the modulating baseband signal. Expressing  $E(t) = E_m \sin \varphi(t)$ ,  $s_a(t)$  is written as

$$s_a(t) = s_1(t) - s_2(t)$$

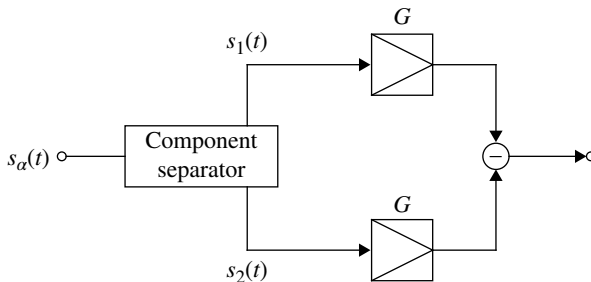
where

$$s_1(t) = \frac{E_m}{2} \sin[\omega_0 t + \theta(t) + \varphi(t)]$$

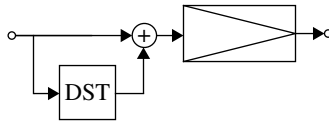
and

$$s_2(t) = \frac{E_m}{2} \sin[\omega_0 t + \theta(t) - \varphi(t)]$$

$s_1(t)$  and  $s_2(t)$  are constant envelope signals, and hence each of them can be amplified through a saturated amplifier with a gain factor of  $G$ . The combined output becomes



**FIGURE 8.24** LINC amplifier.



DST: Distortion generator

**FIGURE 8.25** Linearized power amplifier with predistorter.

$Gs_1(t) - Gs_2(t) = GE_m \sin \varphi(t) \cos[\omega_0 t + \theta(t)] = Gs_a(t)$ . The key issues to the success of this linear power amplifier are the component separation circuit and the balance of the two signal paths.

**8.6.3.4 Predistorter** The predistorter (PD) adds a distorted signal to the input signal in advance to cancel the distortion generated in the amplifiers (Fig. 8.25). Assume the characteristic of an amplifier given by Equation 8.6, and then the predistorted input signal  $x'(t)$  to the amplifier becomes

$$x'(t) = x(t) - \frac{\Delta f_d[x(t)]}{A}$$

where  $x(t)$  denotes the signal to be amplified. The output of the amplifier becomes

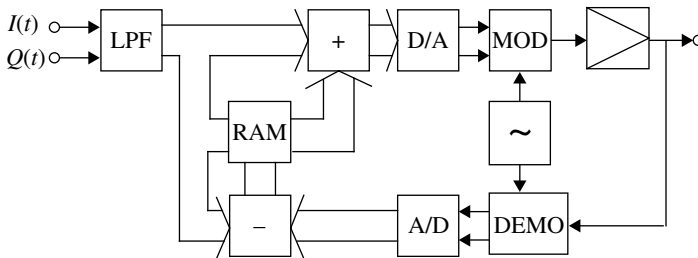
$$y(t) = Ax(t) - \Delta f_d[x(t)] + \Delta f_d\{x(t) - \Delta f_d[x(t)]/A\}$$

When distortion is small ( $|\Delta f_d[x(t)]| \ll A|x(t)|$ ),  $y(t)$  is approximated as

$$y(t) \approx Ax(t)$$

Thus the distortion is canceled.

The advantage of the predistorted amplifier is that, in principle, it is free from the instability problem, since it belongs to an open-loop control. However, the compensating performance degrades if the amplifier parameters deviate from the preset values. In order to resolve this disadvantage, an adaptive control method is proposed by Nagata [19] as shown in Figure 8.26. A feedback path is introduced to detect the error and to automatically correct the predistorter. The predistorter is composed of random access memory (RAM) at baseband frequency. The output of the power amplifier is demodulated and compared with input signal to produce an error signal. The error signal is fed into the predistorter RAM. Thus, the predistorter is adaptive to the change in characteristics of the power amplifier. It should be noted that the bandwidth of the closed loop is much narrower than the bandwidth of the modulating signal; this is not the case for the feedback compensation method. The correction algorithm is based on the iterative method. An algorithm was proposed in Ref. [20] to speed up the convergence time of this method.



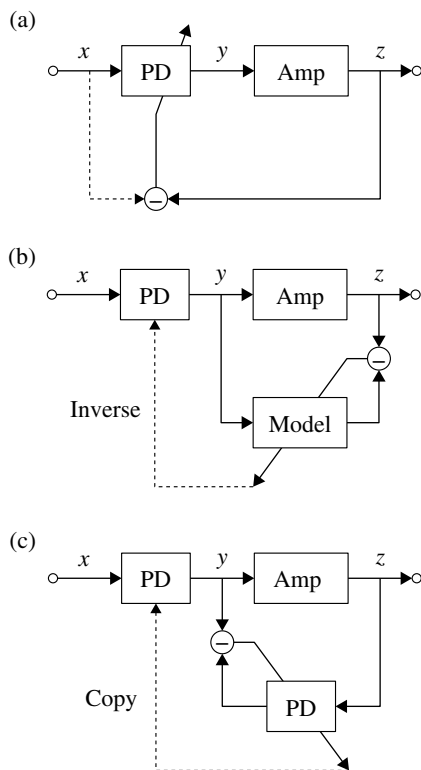
**FIGURE 8.26** Adaptive predistorter amplifier.

The predistorter, especially digital predistorter (DPD) where digital signal processing is used at baseband, has been improved with more sophisticated approaches and is exclusively used in recent base station power amplifiers.

An ideal PD is the inverse circuit for a power amplifier except for the linear gain (Section 2.3.4). However, realization of the ideal PD is difficult since complete estimation of the power amplifier input–output characteristics and an implementation of an ideal inverse circuit are necessary. Therefore, suboptimal PDs are designed based on given conditions. The conditions required for a PD are high compensation performance of nonlinear distortion, low computational complexity, and adaptability to change of the characteristic power amplifier. The adaptability is accomplished by adaptive control of the PD variables on demand generated by monitoring errors defined for the PD control. A measure of compensation performance is the adjacent channel leakage power ratio (ACPR), which should be lower than the given threshold value. As for a measure for in-band distortion, error vector magnitude (EVM), which is defined as the average magnitude of deviation (error) from the modulated signal constellation points, is used. The maximum allowable EVM depends on systems: a lower EVM is required for a higher-level modulation scheme. It is difficult to design a PD with a criterion that satisfies both ACPR and EVM. One of the reasons for this is that we should use a peak to average power ratio (PAPR) (Section 6.5.1) reduction method for efficient operation of PD power amplifiers. The PAPR reduction method, such as clipping and filtering, tends to be incompatible with the aforementioned two requirements and circuit complexity.

We have two kinds of PD design methods. One is an indirect method, where the inverse circuit is given for a power amplifier model whose parameters are determined through approximation of the power amplifier characteristics (Fig. 8.27b). The other is a direct method where PD parameters are given so as to make the remained distortion minimum (Fig. 8.27a). The modified version of the latter is shown in Figure 8.27c: an auxiliary PD is placed at the output of the power amplifier, the PD parameters are determined, and the copy of the PD becomes the actual PD in front of the power amplifier. The principle of the PD implementation is based on the fact that the inverse (nonlinear) circuit may be placed either in front or at the back of a target circuit (Example 2.12). This method has an advantage in that we can make the PD active after the auxiliary PD is run in an offline mode and a stable convergence is confirmed. Furthermore, with this method, coefficients of a power series, which models the power amplifier, can be obtained in a closed form as discussed later.





**FIGURE 8.27** Adaptive predistorters. (a) Direct learning method, (b) estimated and inverse method, and (c) indirect learning method.

**8.6.3.5 Modeling Power Amplifiers** The fact that output of a circuit depends on not only the signal at the present time but also at the past time is called the memory effect. This fact means that the circuit frequency transfer function never satisfies the distortion-free condition (Section 2.3.4). The amplifier model becomes simple if the memory effect is absent. Then output signal  $y(n)$  is determined by input signal  $x(n)$ . We use the equivalent baseband complex expression omitting the time  $n$  as

$$y = G(|x|)e^{j\theta(|x|)}x$$

where gain  $G(|x|)$  and phase  $\theta(|x|)$  are real functions. Functions  $G(|x|)|x|$  and  $\theta(|x|)$  of  $|x|$  are called AM–AM and AM–PM characteristics, respectively. These characteristics are measured with a network analyzer by sweeping the input power. The measured examples are shown in Figure 5.36. As the output power increases to saturated region, the gain decreases, the phase shift becomes large, and the power efficiency increases.

An input–output relation for a narrowband band-pass signal is given with power series expression as (Section 2.1.6)

$$y = a_1 x + a_3 |x|^2 x + a_5 |x|^4 x + \cdots = \left( \sum_{i=0} a_{2i+1} |x|^{2i} \right) x$$

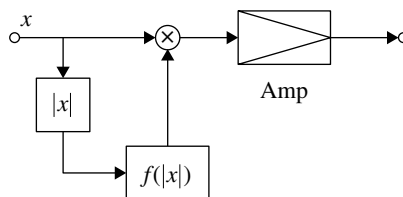
where if the power amplifier has no memory effect for all the higher harmonics (integer multiples of the carrier frequency), phases of coefficients  $a_i$  take the same value. Then the AM–PM characteristics become flat. However, actually, the coefficients  $a_i$  take different phases reflecting the frequency characteristics of a band-pass filter in the power amplifier to block the higher harmonic components. By putting

$$\sum_{i=0} a_{2i+1} |x|^{2i} = G(|x|) e^{j\theta(|x|)}$$

we can obtain the AM–AM and AM–PM characteristics [21].

The principle of PD is to control gain and phase according to the input signal level  $|x|$  so that the gain  $G(|x|)$  and phase  $\theta(|x|)$  characteristics never depend on  $|x|$  (Fig. 8.28). The PD is implemented with a lookup table or an inverse function of the power series  $\sum_{i=0} a_{2i+1} |x|^{2i}$ . The argument is  $|x|$  for both systems. The complex gain  $G(|x|) e^{j\theta(|x|)}$  is a function of  $|x|^2$  and not of odd power of  $|x|$  (Section 2.1.6). It is known, however, nonlinear compensation performance with fixed number of variables is a little improved with use of the odd power components [22].

As a signal bandwidth increases, the transfer function of a power amplifier and the memory effect become not to be ignored [23]. The causes for the memory effect are frequency dependence of RF circuits, dc-power supply circuits, and thermal dissipation mechanism. The thermal effect causes the memory effect through fluctuation of amplifying characteristics corresponding to fluctuation of the junction temperature due to thermal loss power change depending on the input power in a transistor. At an RF band, the frequency transfer function in a bandwidth of input signal is important. Usually, the memory effect is small when the ratio of the signal bandwidth to the center frequency is low. The memory effect due to the dc-power supply circuit is caused by nonlinear distortion of even order as follows: the dc-power supply circuit is designed so as to supply dc power while blocking the RF signals. The inductance  $L$  of a dc-lead line or a choke coil and its impedance  $j\omega L$  are ignored for a low frequency, but are not neglected for a high frequency. The even-order nonlinear distortions appear at baseband and even-order higher harmonics band, although they never appear at the fundamental RF band (Section 2.1.6). The  $2n$ -order distortion causes a signal bandwidth spread by  $2n$  times. The baseband signal caused by these even-order



**FIGURE 8.28** Predistorter implementation (without memory effect).

distortions has higher-frequency components and the impedance  $j\omega L$  looked from a transistor becomes high. Consequently, the dc-bias point of the transistor fluctuates corresponding to the baseband even-order distorted signal voltage, resulting into modulation of the input RF signal and new generation of odd-order distortions, which have the fundamental frequency component. The newly generated odd-order distorted signal reflects the frequency transfer function  $H(\omega)$  of the dc-power supply circuit, which has a strong frequency dependence.

Let us consider a  $2i$ -order distortion. For a complex baseband signal  $x(t)$ , the baseband distorted signal is expressed as  $|x(t)|^2$ . Then the fundamental RF signal is given as

$$y(t) = \left[ b_i |x(t)|^{2i} * h(t) \right] x(t)$$

where  $b_i$  is a constant and  $h(t)$  ( $\leftrightarrow H(\omega)$ ) is the impulse response of the dc-power supply circuit. If a sampled signal notation is used, we have

$$y(n) = \sum_{j=0}^M h(j) \sum_{i=0}^N b_i |x(n-j)|^{2i} x(n)$$

This expression corresponds to the so-called amplitude power series model [24]. If the impedance of the dc circuit is expressed with only an inductance  $L$  and the current is denoted as  $|x(t)|^{2i}$ , the voltage becomes

$$v_i(t) = L \frac{d}{dt} |x(t)|^{2i}$$

Approximating the differentiation by difference, we have

$$y(n) = \sum_{i=0}^N b_i \left[ |x(n)|^{2i} - |x(n-1)|^{2i} \right] x(n)$$

The most general expression of nonlinear distortion of an amplifier including memory effect is the Volterra power series (Section 2.1.6). This is shown with a sampled system expression as

$$y(n) = \sum_{k=1}^K \sum_{i_1=0}^M \cdots \sum_{i_p=0}^M h_p(i_1, i_2, \dots, i_p) \prod_{j=1}^k x(n-i_j) \quad (8.11)$$

where  $h_p(i_1, i_2, \dots, i_p)$  are the model parameters,  $K$  is the highest order of nonlinear distortion, and  $M$  is the memory length. This model is not practical because of a high number of parameters. A simplified version of this is the memory polynomial model [25] expressed as

$$y(n) = \sum_{j=0}^M \sum_{i=1}^N a_{ji} |x(n-j)|^{i-1} x(n-j)$$

This expression corresponds to a matrix expression with only the diagonal elements of the matrix, which is the matrix expression of Equation 8.11. The order of distortion is expanded to include odd orders. This model never expresses the terms  $|x(n-j)|^{2i} x(n)$  (nondiagonal terms) in Equation 8.11, although they are important to express an actual power model with the memory effect. A generalized memory polynomial model is proposed by extending the memory polynomial model [26].

It is a natural approach for us to add a memory circuit to a memoryless circuit, since the memory effect becomes significant as signal bandwidth increases. We have two choices, that is, serial circuit and parallel circuit models (Fig. 8.29). The serial circuit model has two combinations, where any of the two circuits comes first. The inverse circuit for the serial circuit becomes the inverse circuits serially connected in the opposite order of the target circuits. For the parallel circuit model, the inverse circuit is constructed with a feedback circuit, where the feedback part of the circuit consists of distortion components (Fig. 8.30). In this system, output for a feedback circuit without memory should be determined by solving a nonlinear equation, since time recursive approach is not applicable in contrast to solving output for a feedback memory circuit. To solve the nonlinear equation, an iterative method is used for a fixed time as follows: the initial output value is arbitrarily set as  $y_0 = x$  and proceeds

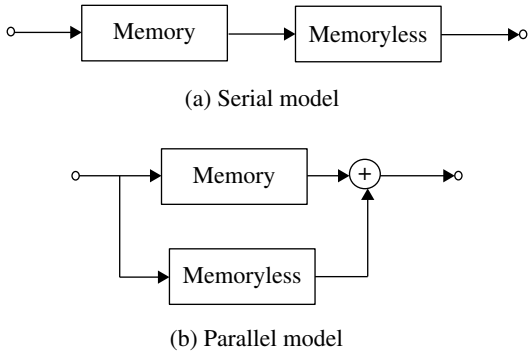


FIGURE 8.29 Amplifier model with memory effect.

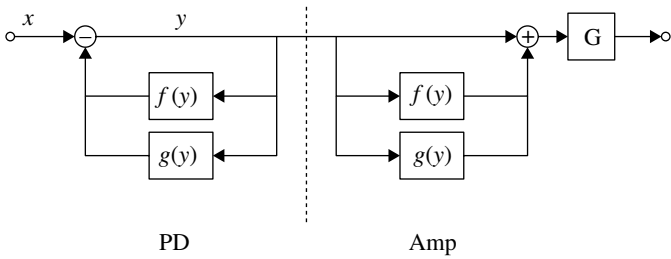


FIGURE 8.30 Predistorter for the parallel-type amplifier model.

as  $y_n = y_{n-1}$  ( $n=1, 2, \dots$ ) till the output  $y_n$  converges. The result converges with a smaller number of iteration for a weaker nonlinear distortion.

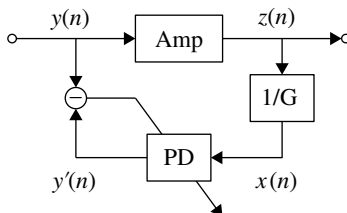
It is practically impossible to completely remove nonlinear distortion with an inverse circuit, since the PD model is not ideal and the circuit and signal processing accuracy are finite. In some occasion, the inverse circuit never exists: for an example of an input signal whose level swings into the saturation level of an amplifier, equalization or linearization is impossible. Therefore, we should implement an inverse circuit approximately with suboptimum parameters given under some evaluation criteria. The MMSE criterion is usually adopted, where the average mean square error is minimized: the error is defined as the difference in ideal and actual outputs. Denoting input signal, output signal, and linear gain by  $x(t)$ ,  $z(t)$ , and  $G$ , respectively, the average mean is defined as

$$J = \left\langle |z(t) - Gx(t)|^2 \right\rangle$$

In using actually the criterion, we modify the input signal as  $x(t - \Delta\tau)$ , where  $\Delta\tau$  is an estimated time delay of the amplifier. The effect of estimation error of  $\Delta\tau$  becomes significant as the signal bandwidth increases. To avoid difficulty of the time delay estimation, a method where the error is defined as the out-of-band frequency components, which are generated by nonlinear distortion [27]. A defect of this method is low capability of in-band distortion reduction.

Several algorithms are known to determine the optimum parameters based on the MMSE criterion. A general approach is the perturbation method, where parameters are updated with the trial-and-error technique for a small variation of parameters. The steepest decent method is an iterative technique that uses differential coefficients to reach the optimum parameters (Section 2.5.2). Here, we develop a method that uses a property of the inverse circuit [28]. This method assumes a memory polynomial PD with indirect learning as shown in Figure 8.31. The input  $y(n)$  to the amplifier and the output  $\hat{y}(n)$  of the PD become the same for an ideal situation, since an inverse circuit (PD) may be placed at the output of the target circuit (amplifier) (Section 2.3.4). Here, the input of the PD is normalized considering the linear gain  $G$ . We use the memory polynomial to express the PD input–output characteristics as

$$\hat{y}(n) = \sum_{k=1}^K \sum_{q=0}^Q a_{kq} u_{kq}(n)$$



**FIGURE 8.31** Control parameter determination method.

where  $u_{qn}(n) = |x(n-q)|^{k-1} x(n-q)$ ,  $x(n) = z(n)/G$ . Since  $\hat{y}(n)$  is a linear function of variables  $a_{kq}$ , the optimum parameters on the MMSE criterion are given by linear operation as follows. Consider that the input signals  $y(n)$  and output signals  $z(n)$  ( $n = 1, 2, \dots, N$ ) are obtained, and then the average mean square error is expressed as

$$\begin{aligned} J &= \sum_{n=1}^N |y(n) - \hat{y}(n)|^2 \\ &= \sum_{n=1}^N \left| y(n) - \sum_{k=1}^K \sum_{q=0}^Q a_{kq} u_{kq}(n) \right|^2 \end{aligned}$$

Differentiating this with respect to  $a_{lm}^*$ , we have

$$\begin{aligned} \frac{\partial J}{\partial a_{lm}^*} &= \sum_{n=1}^N \left[ y(n) - \sum_{k=1}^K \sum_{q=0}^Q a_{kq} u_{kq}(n) \right] (-u_{lm}^*) \\ &\quad (l = 1, 2, \dots, K, \quad m = 0, 1, \dots, Q) \end{aligned}$$

Letting  $\partial J / \partial a_{lm}^* = 0$ , we get the following  $K(Q+1)$  linear simultaneous equations

$$\sum_{k=1}^K \sum_{q=0}^Q c_{kqlm} a_{kq} = d_{lm}$$

where  $c_{kqlm} = \sum_{n=1}^N u_{kq}(n) u_{lm}^*(n)$  and  $d_{lm} = \sum_{n=1}^N y(n) u_{lm}^*(n)$  are understood.

By solving the aforementioned equations, the optimum parameters are obtained.

Instead of solving the equations, the minimum mean square algorithm (LMS: Section 2.5.2) is applicable as an iterative method as

$$\begin{aligned} a_{kq}(n+1) &= a_{kq}(n) - \mu \frac{\partial}{\partial a_{lm}^*} J(n) \\ &= a_{kq}(n) + \mu \varepsilon(n) u_{lm}^*(n) \quad (n = 0, 1, \dots) \end{aligned}$$

where  $\varepsilon(n) = y(n) - \hat{y}(n)$  and  $\mu$  is a small constant. Although the memory polynomial is used in the aforementioned discussion, we can add any terms, for example,  $b_{qi} |y(n-q)|^{i-1} y(n)$ , ( $b_{qi}$ : new variables), since power series  $u_{kq}(n)$  are arbitrary.

*Orthogonal Series Expansion:* To enhance the convergent speed of iterative calculation in the LMS algorithm, a method with orthogonal series is known [29]. We assume that an output  $y$  is given with power series of  $x$ . The orthogonal series  $\varphi_k(x)$  is defined as

$$\begin{aligned} \langle \varphi_k(x) \varphi_l^*(x) \rangle &= \sum_{n=1}^N \varphi_k(x(n)) \varphi_l^*(x(n)) \\ &= \delta_{kl} \end{aligned}$$

where  $\delta_{kl} = \begin{cases} 1 & (k = l) \\ 0 & (k \neq l) \end{cases}$ .

We express  $\varphi_k(x)$  as

$$\varphi_k(x) = \sum_{i=0}^k c_{ki} u_i(x)$$

where  $u_i(x)$  are power series of  $x$  and expressed as  $u_i(x) = |x|^{2i} x$  for an example. If we take expression  $y = \sum_{k=0}^N b_k \varphi_k(x)$  and  $\hat{y} = \sum_{k=0}^N \hat{b}_k \varphi_k(x)$  instead of  $y = \sum_{k=0}^N a_k u_k(x)$  and  $\hat{y} = \sum_{k=0}^N \hat{a}_k u_k(x)$  respectively, the updating value  $\Delta b_k(n) = \mu[y(n) - \hat{y}(n)] \varphi_k^*(x(n))$  for coefficients  $b_k(n)$  becomes stochastically independent of the other updating values  $\Delta b_l(n)$  ( $l \neq k$ ), and therefore the convergent speed increases.

The coefficients  $c_{ki}$  in  $\varphi_k(x) = \sum_{i=0}^k c_{ki} u_i(x)$  can be determined successively starting from the initial value  $c_{10}$  using the orthogonal series relations.

**8.6.3.6 PAPR Reduction** The PAPR of an input signal affects the compatibility of power efficiency and nonlinear distortion in a power amplifier. To keep the nonlinear distortion below a given level, the average power should be more reduced (backed off) from the saturation level as the PAPR increases, and consequently the power efficiency drops. From this viewpoint, modulation schemes with a low PAPR such as offset QPSK and HPSK are recommended (Section 6.3). Even with such a modulation method, the PAPR increases if preequalization (or filtering; Section 7.3.6) is introduced at the transmit side. The PAPR increases irrespective of modulation schemes if multiple signals are added (or multiplexed) at the same time such as OFDM and CDM (Section 3.7). Here, PAPR reduction methods [30] are described assuming an OFDM transmission system.

In reducing PAPR, the peak power should be decreased since the average power is fixed to satisfy the required received signal to noise power ratio. The first approach for the PAPR reduction is to control the phases of each multiplexed signal so as not to take the same phase: the highest peak power is generated when all the signals take the same phase at the same time. To demodulate the phase-controlled signal at receive side, side information to compensate the controlled phase should be sent along with the transmit signal.

A fine control of phases of each (subcarrier) signal results in increase of computational complexity for optimum phase search and in decreases of channel efficiency due to the increased side information. To mitigate the difficulty, a method called selective mapping is known: the phase control is carried out selecting the best one of the phase control patterns that are representatively prepared beforehand. An index for the pattern is sent as the side information. The number of patterns increases as the number of subcarrier increases. To cope with this difficulty, a method called partial transmit sequence is known where subcarriers are divided into groups and signals within a group are subjected to the same phase control.

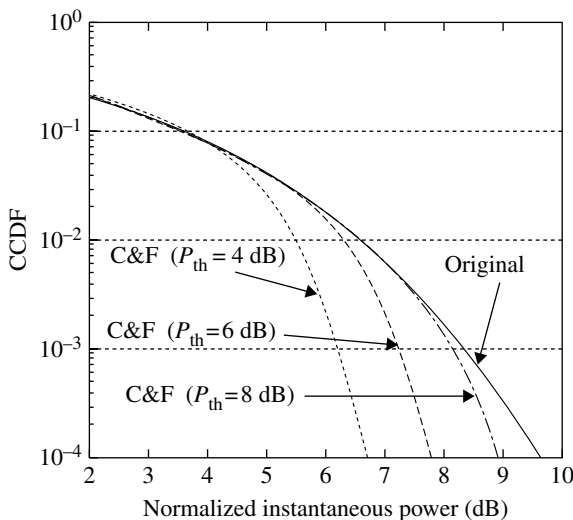
Another method [31] is to employ auxiliary subcarriers at given frequencies to suppress power peaks. This method never requires to send side information.

The aforementioned methods never cause signal distortion. In contrast to those methods, the peak clipping and filtering method generates nonlinear distortion. In this method, the signal level is clipped to suppress the signal peaks under a given

value and applied to a band-limiting filter. Out-of-band frequency components are produced through the nonlinear operation, namely, clipping. The band-limiting filter is used to suppress the out-of-band components. The peak levels may grow over the threshold through the band limit filtering, and this process is repeated for several times. The band-limiting process never suppresses the in-band-distortion or modulation error defined such as EVM. This method is the most practical one in a sense that it never requires side information and standardization for system compatibility and its computational complexity is comparatively small. The CCDF (Section 6.3.1) characteristics with this method are shown in Figure 8.32. The degree of peak limit should be decided taking the deterioration of EVM: the EVM increases as the peak limit threshold becomes low.

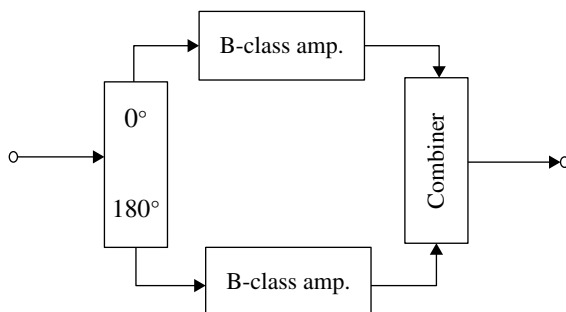
**8.6.3.7 The Doherty Power Amplifier** Although nonlinearity compensation and PAPR reduction are effective for compatibility between high linearity and power efficiency in a power amplifier, the power efficiency improvement is small when signal power is low. In such a case, the power efficiency of the amplifier itself should be improved. The most useful method is the amplifier invented by W.H. Doherty in 1936 [32]. Combined with PD using modern signal processing technique, this amplifier is commonly applied in cellular base stations. Here, we discuss it briefly.

To understand the principle of the Doherty amplifier, it is useful for us to compare it with the push–pull amplifier, which had been known earlier than the Doherty amplifier. Doherty himself might get his new idea through this thinking approach. The push–pull amplifier is shown in Figure 8.33. The input signal is divided into two



**FIGURE 8.32** An example result on peak to average power ratio suppression (128 carrier OFDM signal).

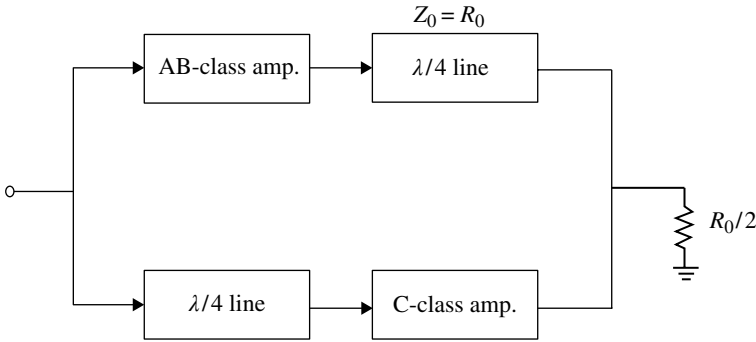




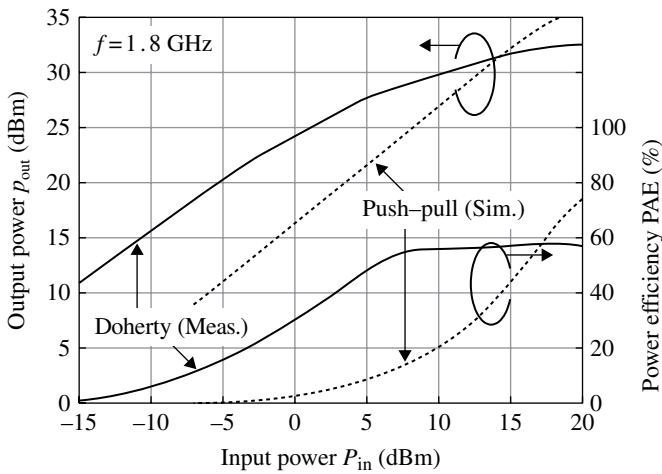
**FIGURE 8.33** The balanced (push-pull) amplifier.

signals with opposite phases, and then each signal is applied to a class B amplifier. One of the class B amplifiers whose input is positive becomes active, because of class B operation of the amplifier. The amplifier's two output signals are serially combined. A high power efficiency owing to class B operation and a high linearity due to parallel amplifiers are obtained. When signal level is low, however, the power efficiency drops since both amplifiers are in class B operation. To solve the problem, let us modify the push-pull amplifier so that one of the amplifiers is operated in class AB mode and the other is operated in class C mode. Then class AB amplifier shows a high power efficiency for a small signal since it becomes saturated. Class C amplifier becomes active when the input signal level is high. The combined signal gives an almost the same saturated power as a push-pull amplifier. A defect of the modified amplifier is that it shows high nonlinear distortion: when input signal power increases to a level that class C amplifier begins to be active, the combined output power suddenly increases, and consequently, the input-output power linear relation (AM-AM) is destroyed. In another explanation, the total gain of the amplifier jumps at the input level.

Figure 8.34 shows the Doherty amplifier. As shown in Figure 8.35, it gives a high power efficiency even for a low signal level, since class AB (carrier) and class C (peak) amplifiers are operated in parallel. In addition to this, the input and output power shows a fairly linear relation. The Doherty amplifier prevents from the linearity degradation by using an impedance transformers (quarter-wavelength transmission line) to combine the output signals: the carrier (class AB) amplifier's gain drops as the peak amplifier (class C) becomes active to make the total gain remain the same. The mechanism is explained in the following. We assume the load impedance of the Doherty amplifier to be  $R_0/2$ . The output of the carrier amplifier is connected via a quarter-wavelength ( $\lambda/4$ ) line with a characteristic impedance of  $R_0$  to the load, and the output of the peak amplifier is connected to the load in parallel with the ( $\lambda/4$ ) line. When input signal power is low, the impedance looking into the output of the peak amplifier becomes infinite in principle, since the peak amplifier is not active for such a low-level input signal. Then the ( $\lambda/4$ ) line is equivalently terminated with a load of  $R_0/2$ . In this case, the impedance looking from the output of the carrier amplifier into the ( $\lambda/4$ ) line becomes, from a property of a ( $\lambda/4$ ) line (Appendix 8.A),



**FIGURE 8.34** The Doherty amplifier.



**FIGURE 8.35** An example of output power and power efficiency versus input power for a Doherty power amplifier. The performance with a push-pull amplifier is also shown. Courtesy of Prof. Honjyo of the University of Electro-Communications.

$$Z_{in} = \frac{R_0^2}{R_L (= R_0/2)} = 2R_0$$

where  $R_L$  is the terminating impedance of the  $\lambda/4$  line. The equivalent load impedance of the carrier amplifier is  $2R_0$ .

Next, we consider a case when the peak amplifier becomes active. If the output currents of the carrier and the peak amplifiers are the same including their phases (a  $\lambda/4$  line is connected to the input of the peak amplifier to adjust the phase), the impedance looking from the output of the  $\lambda/4$  line and the peak amplifier becomes  $R_0$  considering the currents contributed from the counterpart. Then the impedance looking from the output of the carrier amplifier becomes  $Z_{in} = R_0^2/R_0 = R_0$ . When the load impedance

decreases from  $2R_0$  to  $R_0$ , the carrier amplifier's output power, and therefore its gain, drops to be one half, since the transistor in the amplifier is approximated as a current source. Thus, the gain of the carrier amplifier is automatically controlled with respect to the peak amplifier's mode of operation, and consequently, the gain versus input power becomes flat over a wide range of input level, that is, the linearity is improved.

The linearity of the Doherty amplifier is not sufficient, and a nonlinearity compensation method is required for actual applications. In the Doherty paper [32], a feedback method was adopted since the amplifier was applied for an analog AM broadcasting system, whose bandwidth is narrow. Figure 8.36 shows performance of the commercial Doherty amplifiers with a DPD for application to cellular base stations.

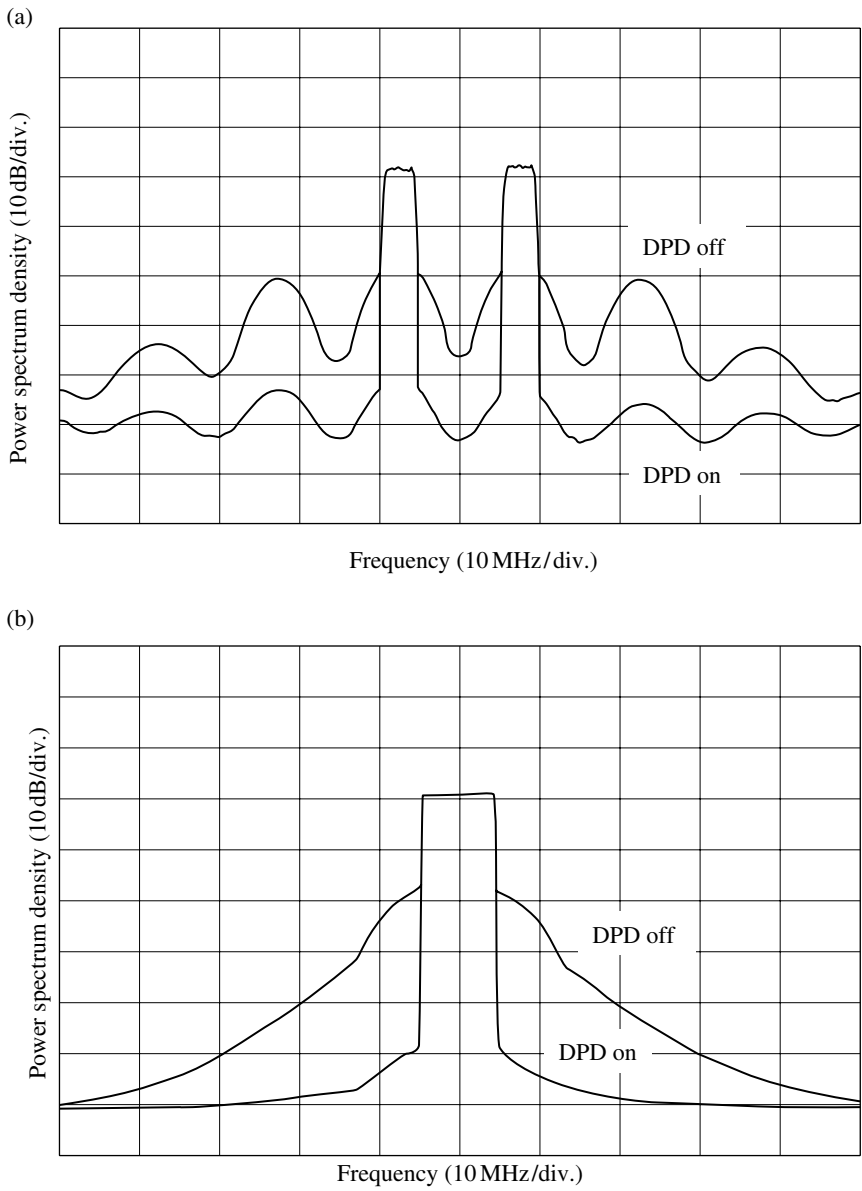
#### 8.6.4 Transmit Power Control

Transmit power control reduces the interference with other signals by preventing signal transmission at unnecessary power levels. The difficulty of circuit implementation for the transmit power control differs between the types of transmit power amplifiers employed: highly saturated (e.g., class C power) amplifiers or (quasi) linear amplifiers. The input–output relations of class C and class AB amplifiers are shown in Figure 8.37. We can see for class C power amplifier that it is hard to manage the output power levels, since a very small change in the input signal level corresponds to a big change in the output power; we must introduce a feedback control technique. On the other hand, for a quasilinear power amplifier, we can merely use a variable attenuator at the input of the power amplifier (feed-forward control).

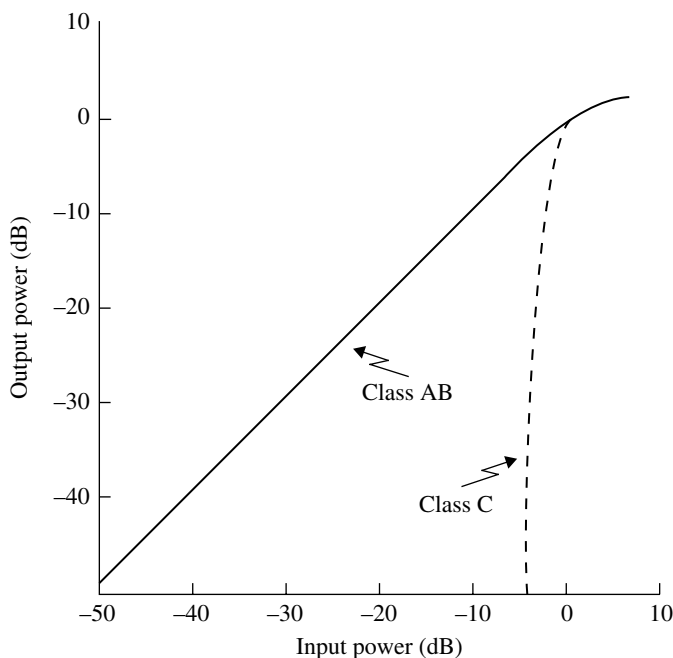
Transmit power control is necessary for burst signal transmission in a TDMA system. The output signal level should be gradually ramped up and down at the beginning and ending of the burst signal to prevent the spectrum from spreading at these times. Figure 8.38 shows a power amplifier with output signal waveform control for the aforementioned purpose. Class C amplifier is assumed. Such feedback control is unnecessary for a linear power amplifier. Hence, to avoid the feedback control circuit, a (quasi)linear power amplifier is often adopted in TDMA systems, even when constant envelope modulation is used: the use of a linear power amplifier diminishes the advantage of constant envelope modulation described in Section 5.2.

The transmit power control should be implemented automatically, corresponding to fading channels. Feed-forward and feedback methods that achieve this result are known. The former is used for a system where the uplink and downlink are correlated such as in the TDD system using one frequency. The latter is used for a two-frequency duplex system where no correlation is expected between the downlink and uplink. In the feedback control method, received signal information at the receiver is reported to the transmitter.

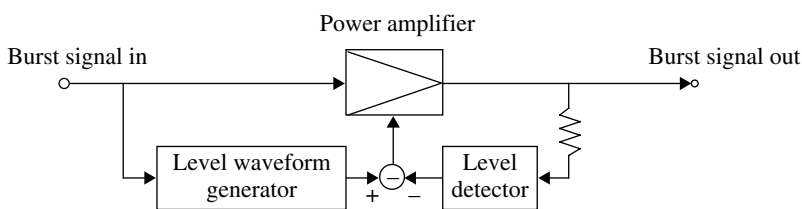
In a fast-fading channel, the tracking speed of the automatic power control system is important. Prediction of signal level using past data is effective for this purpose. In predicting a signal level, prediction based on signal amplitude is not effective, since the amplitude is subjected to an acute notchlike fading characteristic. Instead, prediction based on inphase and quadrature components of the received signal is more effective [33], since the signal components will never have such fading notches.



**FIGURE 8.36** Performance of the GaN Transistor Doherty power amplifier with digital predistorter. Frequency band, 2GHz; transmit power, 40W; final-stage power efficiency, 50–60%. (a) WCDMA signal (15MHz spaced two carriers, bandwidth: 20MHz), (b) LTE signal (bandwidth: 10MHz) (Courtesy of Hitachi Kokusai Co. Ltd.).



**FIGURE 8.37** Input-output relation of class AB and class C amplifiers.



**FIGURE 8.38** Output signal-level waveform-controlled amplifier.

Different criteria for power control are known: one is to keep the received signal level constant, and the other is to keep signal to interference power ratio at a given value. For the latter, a distributed controlled system is analyzed in Ref. [34]. An automatic power control system with the latter criterion is proposed in Ref. [35] to make the system stable.

## 8.7 RECEIVER CIRCUITS

### 8.7.1 AGC Circuit

An AGC circuit is needed to cope with the wide dynamic range of the received signal in a mobile radio channel. The constant envelope modulation signal can be received with an amplitude limiter instead of an AGC circuit. Even with

this kind of modulation, AGC is required if a channel equalizer is used since channel equalization is difficult if the distorted signal is amplitude limited.

A typical AGC circuit is shown in Figure 8.39. A pilot signal may be used to allow level detection to be free from interference from signal-level perturbations by modulation. The output signal level is feedback controlled to a constant value. The bandwidth of the LPF must be narrow so as not to track the modulated signal amplitude. In this system, tracking speed versus stability of the operation becomes an issue.

A feed-forward AGC is shown in Figure 8.40. The average level is detected and the inverse of it is multiplied with the input signal to give a constant-level signal preserving the amplitude waveform of the modulated signal. Another version of feed-forward AGC is shown in Figure 8.41. The received signal is amplitude limited and amplitude detected in parallel. The amplitude detector deals only with the signal amplitude due to modulation. The constant-level signal with a constant average level is restored at the output of the multiplier.

The circuit shown in Figure 8.42 [36] is not the AGC circuit; however, it resolves the wide dynamic range problem. The input signal is applied to a log amplifier, whose typical implementation is shown in Figure 8.43. The log amplifier outputs the logarithm of the input signal level or received signal strength indicator (RSSI) and amplitude-limited signal. The amplitude-limited signal is fed to a phase detector. The

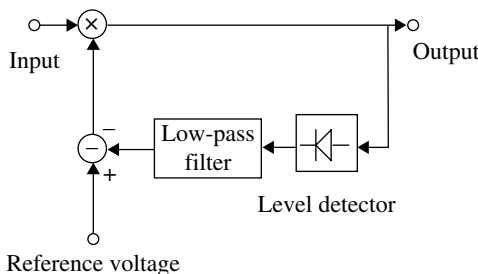


FIGURE 8.39 Feedback-controlled automatic gain control circuit.

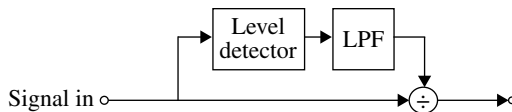


FIGURE 8.40 Feed-forward automatic gain control.

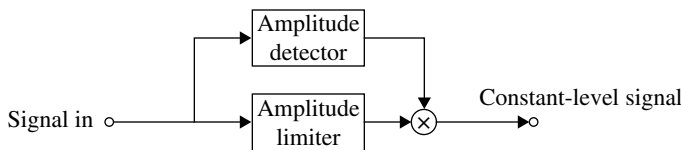
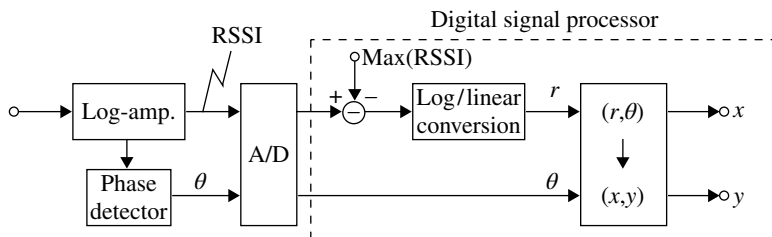
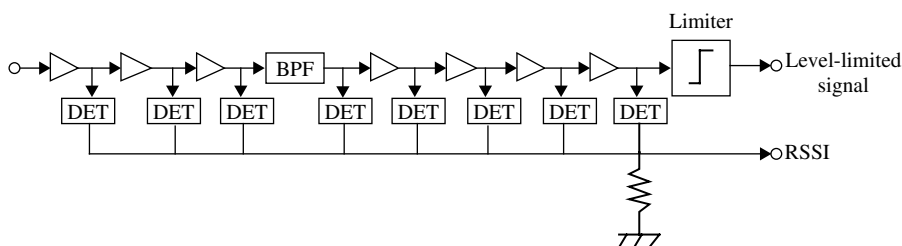


FIGURE 8.41 Feed-forward automatic gain control with amplitude limiter.



**FIGURE 8.42** Digital signal processing receiver coping with wide dynamic range signal [36]. Reproduced by permission of Copyright © 1994, IEICE.



**FIGURE 8.43** Log amplifier with received signal strength indicator (RSSI) [36]. Reproduced by permission of Copyright © 1994, IEICE.)

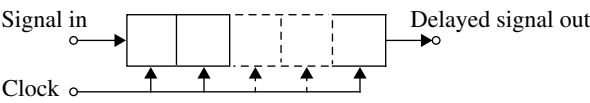
detected phase  $\theta$  and the RSSI signal are applied to the analog-to-digital converter. Since the signal level is compressed by the log amplifier, the dynamic range of the analog-to-digital converter never comes in question. The maximum value of sampled RSSI signal in a block of the samples is found and is subtracted from the samples to normalize the received signal level. After this, the RSSI signal is log to linear converted. The amplitude and phase are translated into quadrature components for further signal processing such as demodulation.

### 8.7.2 Signal Processing with Logic Circuits

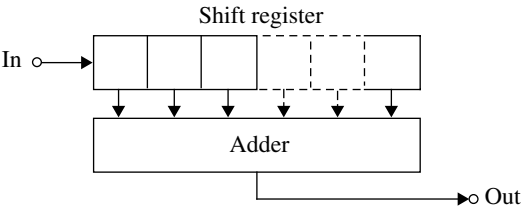
In this system, baseband pulse signals and angle-modulated signals are processed with logic circuits after conversion to two-level logic signals.

A delay line is realized with a shift register as shown in Figure 8.44. The time delay is accurate since it is controlled with a clock frequency and the number of shift register stages. The size of the shift register becomes large when the signal frequency is high or the required time delay is long.

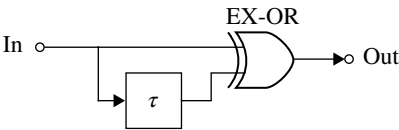
Figure 8.45 shows a circuit that takes the moving average of the input signal. The output is obtained as an integer and is suitable for treatment by a microprocessor thereafter. The moving average is equivalent to an LPF with transfer function  $H(\omega) = \sin(\omega T)/\omega T$ , where  $T$  is the duration of the average and is equivalent to integrate-and-dump filtering at the sampling instant (Section 2.3.2).



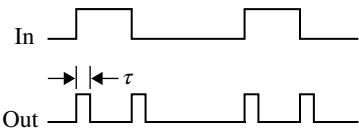
**FIGURE 8.44** Delay line consisting of shift register.



**FIGURE 8.45** A moving-average circuit.



(a) Circuit



(b) Waveform

**FIGURE 8.46** Time-differentiating and rectifying circuit.

A time-differentiating and rectifying circuit is shown in Figure 8.46a with schematic explanation of its operation (Fig. 8.46b). This circuit is useful in a clock recovery system.

A phase detector or comparator is shown in Figure 8.47 together with the exclusive-OR (EX-OR) output waveform. The moving-average circuit in Figure 8.45 may be used for the LPF. Another method for the LPF is a counter that counts the number of fast pulses during the high-level state of the EX-OR output signal for a period of the local frequency. The counted number is proportional to the phase difference between the signals. The phase detection characteristic curve is linear. This phase detector covers a phase difference of 0 to  $\pi$ . This circuit works as a frequency downconverter when the LPF cutoff frequency is properly selected. The quadrature phase detector is shown in Figure 8.48. This circuit covers a phase difference of  $-\pi$  to  $\pi$ .



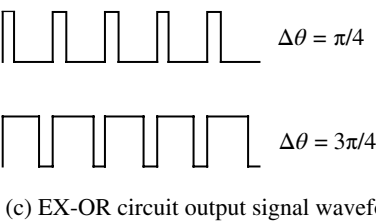
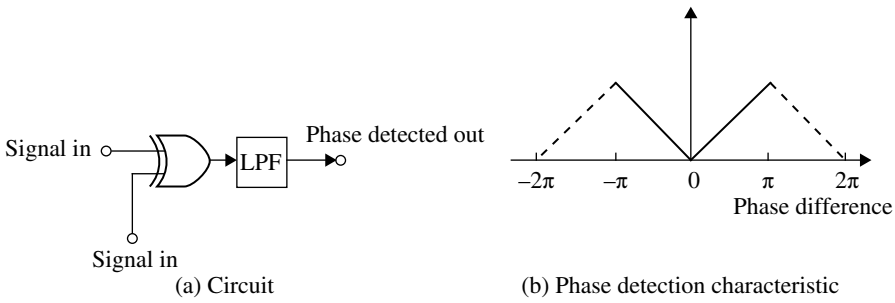


FIGURE 8.47 Phase detector.

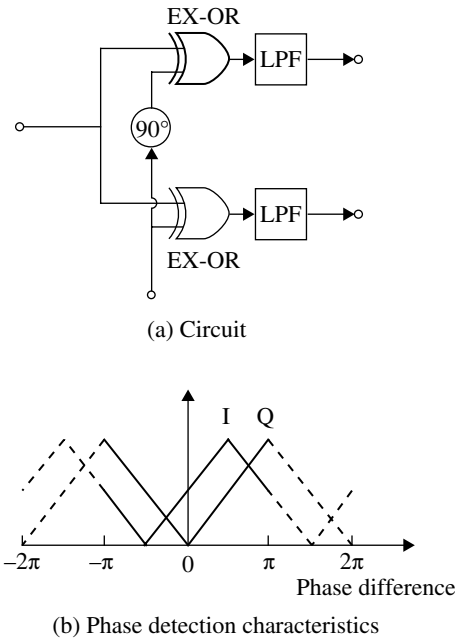


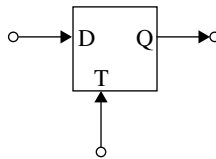
FIGURE 8.48 Quadrature phase detector.

A phase detector that detects only two-state phase differences, that is, advance or delay, is shown in Figure 8.49. This is a D-type flip-flop circuit. We need no LPF for this phase detector.

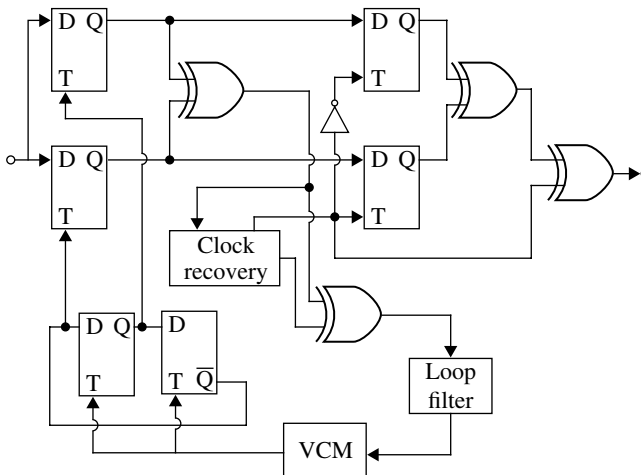
### 8.7.3 Demodulator

A coherent demodulator for MSK, GMSK, and tamed FM is shown in Figure 8.50 [37]. It is assumed that the input signal is band-pass filtered and converted into a logic signal. Most of the circuits consist of logic circuits. Operation of this circuit may be understood by referring to previously described circuit elements and comparing this circuit with that in Figure 6.11.

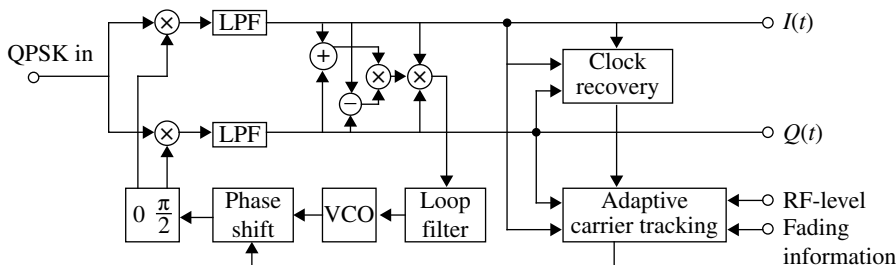
Coherent demodulators under fast-fading conditions have poorer floor error rate performance than noncoherent demodulators, for example, differential demodulators, although they give better performance under static or slow-fading conditions. A coherent demodulator that improves performance under fast fading is shown in Figure 8.51 [38]. The demodulator works with dual modes of operation: when the fading frequency is slow and the RF signal level is low, it selects the conventional Costas loop carrier recovery system; otherwise, it selects the adaptive carrier tracking



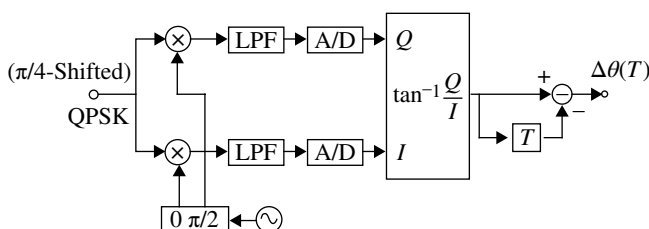
**FIGURE 8.49** Two-state phase detector.



**FIGURE 8.50** A coherent demodulator for MSK, GMSK, and tamed FM [37]. Reproduced by permission of Copyright © 1981, IEEE.



**FIGURE 8.51** Block diagram of dual-mode carrier recovery system with dual mode of operation with Costas loop and adaptive carrier tracking.



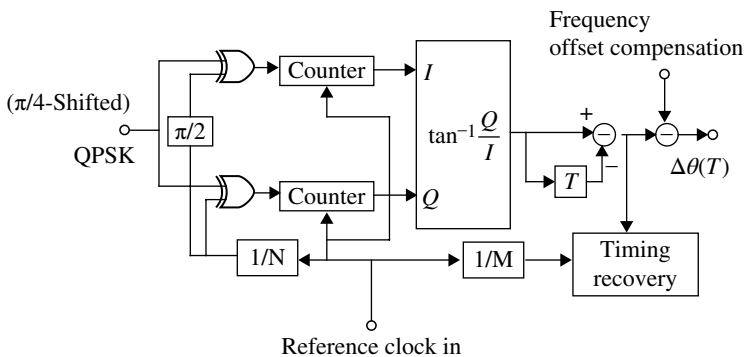
**FIGURE 8.52** Differential detector of  $(\pi/4)$  shifted QPSK.

mode. In the adaptive carrier tracking mode, the carrier phase deviation due to fast fading is corrected if the deviation becomes larger than a given threshold value. It is shown that the adaptive carrier tracking mode of operation is equivalent to differential detection. It therefore shows improved floor error rate performance under fast-fading conditions.

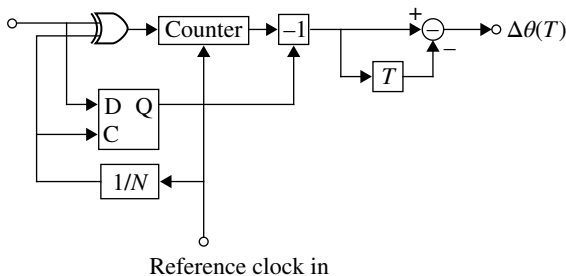
A differential demodulator for QPSK or  $\pi/4$  shifted QPSK can be constructed with a delay line and phase detectors at IF band. Instead of using the delay line, there is another method of detecting instantaneous phase relative to the asynchronous local carrier signal phase. This method is shown in Figure 8.52. The logic circuit-implemented demodulator is shown in Figure 8.53. The quadrature demodulator consists of Ex-OR circuits and counters. From the phase detection characteristics in Figure 8.48, the instantaneous phase  $\theta$  can be found from the absolute value of  $\theta$  and its sign. A differential detector based on this fact is shown in Figure 8.54. The D-type flip-flop circuit detects the sign of the phase difference (i.e., two-level phase) as described before. An all-digital differential detector including postdetection diversity is described in Ref. [39].

The carrier frequency offset worsens the error rate performance of the differential detector. The frequency offset  $\Delta\omega$  results in an extra phase shift  $\Delta\omega T$  in the demodulated signal  $\Delta\theta(T)$ . Thus, the phase error from signal phase shifts of  $\pm\pi/4$  or  $3\pi/4$  for  $\pi/4$  shifted QPSK is used to compensate for the frequency offset [40].

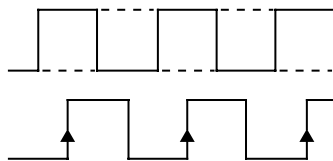
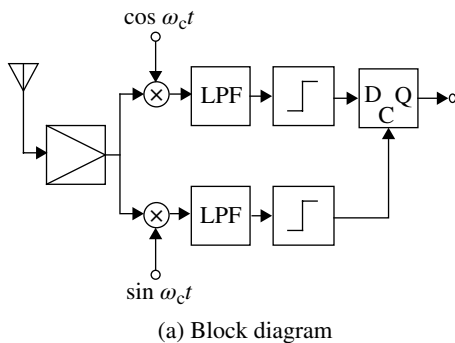
A direct conversion demodulator for two-level FM can be constructed as shown in Figure 8.55 [41]. Limiters and a D-type-flip-flop are used for demodulation. The principle of this demodulation is to detect the direction of rotation of the signal trajectory in



**FIGURE 8.53** Logic circuit-implemented differential detector of  $(\pi/4)$  shifted QPSK.

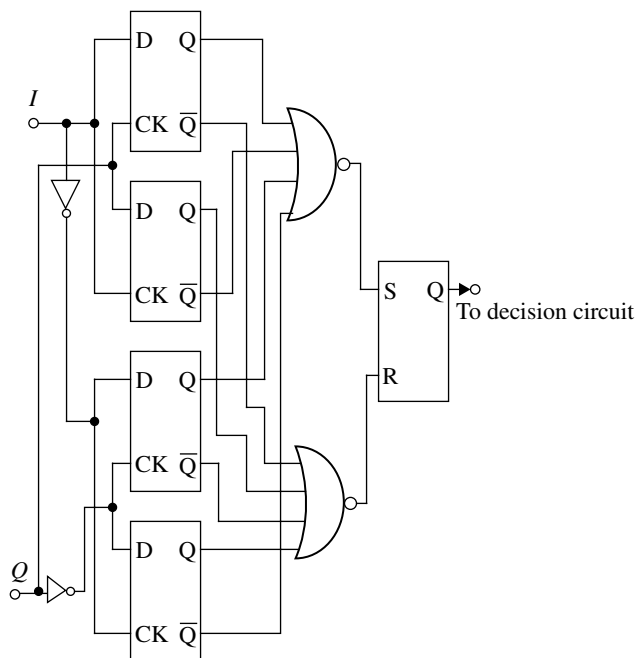


**FIGURE 8.54** Differential detector using an EX-OR phase detector and two-level phase detector.



(b) Waveform

**FIGURE 8.55** Direct conversion demodulator for two-level FM.



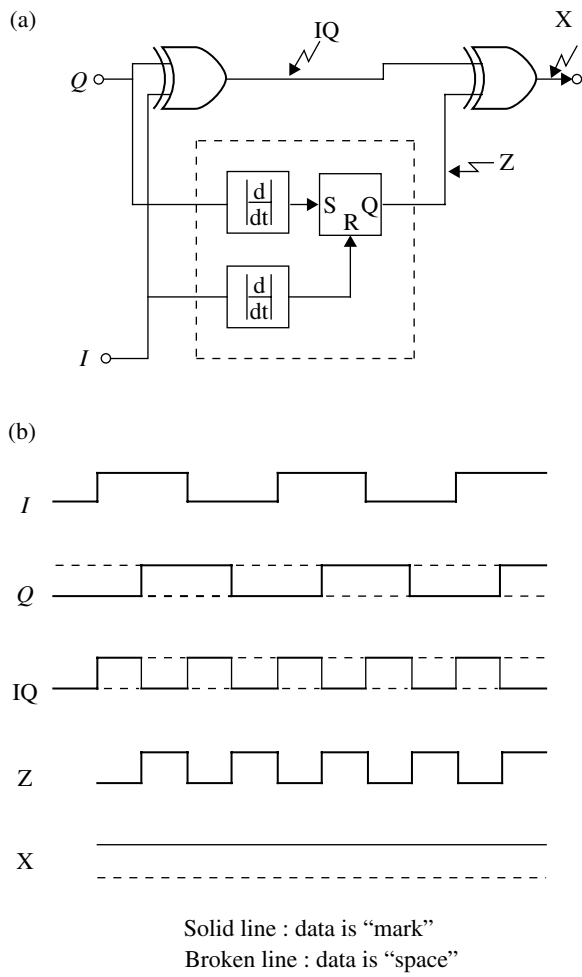
**FIGURE 8.56** FSK demodulator.

the quadrature plane. The detection is made once per rotation of  $2\pi$ . The detector cannot cope with signals with modulation index less than 2. Detection is made four times per  $2\pi$  rotation with the demodulator shown in Figure 8.56 [42]. Detection is made at each rise up and down instant of the quadrature signals. Another demodulator is shown in Figure 8.57 [43]. The principle of this circuit is detection of the rotating direction of the equivalent baseband frequency-doubled signal. Figure 8.58 shows another demodulator [44]. This circuit also operates on the frequency-doubled concept.

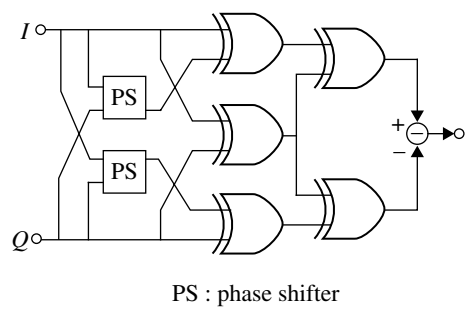
The group modulator or its counterpart, the group demodulator, modulates (demodulates) frequency division multiplexed signals (Fig. 8.59). It modulates/demodulates multiple signals at the same time. Thus, it becomes useful for a multi-carrier transmission system or a base station transceiver for an FDMA system. Group modulators and demodulators using the discrete Fourier transform technique are already described in Sections 6.5.2 and 6.5.3, respectively.

## 8.8 COUNTERMEASURES AGAINST DC BLOCKING AND DC OFFSET

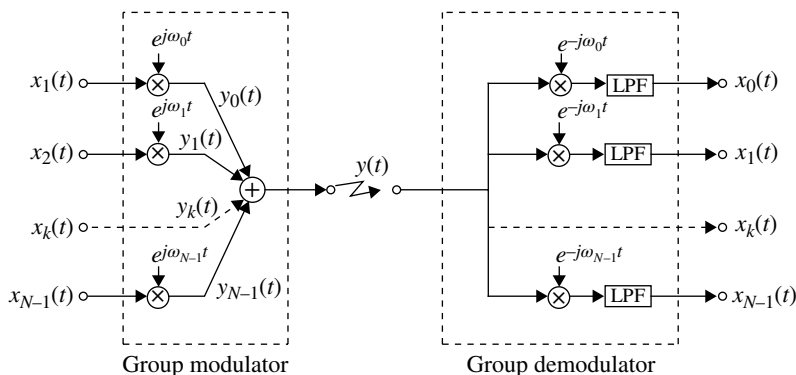
dc blocking or ac coupling is desirable for preventing baseband signals from the dc offsets of the circuits. For signal transmissions through dc-blocked channels, waveforms without dc components are transmitted; for example, the Manchester codes (Section 3.2.4) are used, sacrificing spectrum performance.



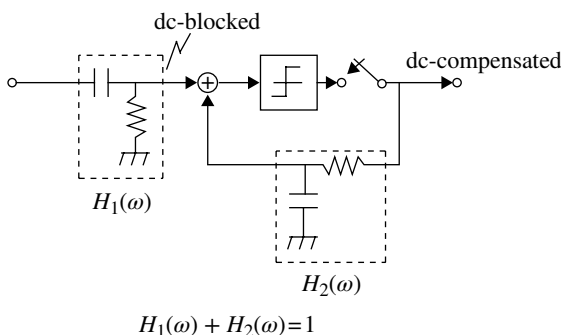
**FIGURE 8.57** FSK demodulator. (a) Block diagram and (b) waveform.



**FIGURE 8.58** FSK demodulator with phase shifter.



**FIGURE 8.59** Group modulator and group demodulator.



**FIGURE 8.60** dc-compensating circuit for digital signal.

Another approach is to use compensating techniques for waveforms distorted by dc-blocked channels. Figure 8.60 shows such a system for digital signal transmission. The lost dc component is restored by applying the received data signals to an LPF. It constitutes a decision feedback equalizer (Section 7.3.3).

Another class of countermeasures against dc offsets is based on detecting the dc offset and subtracting it from the input dc-offset signal as shown in Figure 8.61. The integrator can be replaced with an LPF. Such a technique can be applied to the automatic-level control (ALC) circuit as shown in Figure 8.62. The combined dc- and level-offset control was implemented in a four-level digital FM with limiter-discriminator system [45].

At a base station receiver for a TDMA system, dc-offset control is difficult, since it must be performed for each burst with different dc-offset values. A TDMA digital FM/limiter-discriminator detection system has such a problem, due to carrier frequency difference between the bursts transmitted from subscriber stations: the carrier frequency difference results in a dc offset. As a countermeasure against this problem, a method was proposed in Ref. [46] in which demodulated signals are subtracted from the signal delayed by one symbol duration. By means of this subtraction, the dc offset

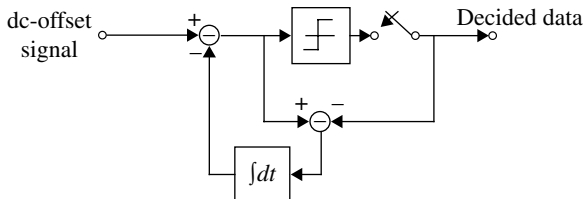


FIGURE 8.61 dc-offset control circuit.

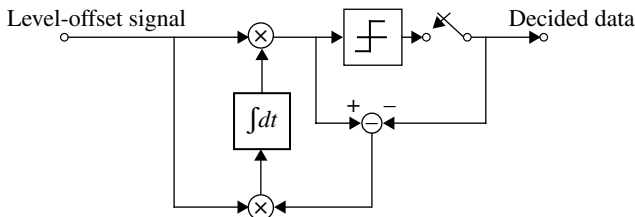


FIGURE 8.62 Automatic-level control circuit.

is removed at the symbol rate. This is equivalent to the partial-response system with an impulse response  $h(t) = \delta(t) - \delta(t - T_s)$ , where  $T_s$  is the symbol duration. However, the error rate performance deteriorates with this method. The performance deterioration is recovered [47] using maximum likelihood sequence estimation (Viterbi algorithm; Section 3.3.8).

For a burst signal receiver, dc- and level-offset compensation in Figures 8.60 and 8.61 is operated using some given preamble bits [48].

The most effective method for dc offset as well as imbalance problem in quadrature demodulation is to use digital signal processing at an IF band: the analog signal is converted into digital signal, and the signal is processed with digital circuits later on in the receiver. The same method is useful for transmitter: baseband signal is processed and upconverted into an IF signal with digital signal processing and then converted into analog signal.

## APPENDIX 8.A QUARTER-WAVELENGTH LINE

The voltage  $V(x)$  and current  $I(x)$  at a position  $x$  in a transmission line are described by the following equation:

$$V(x) = V_i e^{-j\beta x} + V_r e^{j\beta x}, \quad I(x) = \frac{1}{Z_0} (V_i e^{-j\beta x} - V_r e^{j\beta x}) \quad (8.A.1)$$

where the propagation constant  $\beta = 2\pi/\lambda$  ( $\lambda$ : wave length) and  $Z_0$  is called characteristic impedance. If we use a time factor as  $e^{j\omega t}$ , then,  $e^{\pm j\beta x} e^{j\omega t} = e^{j(\pm\beta x + \omega t)}$  and therefore  $V_i e^{-j\beta x}$  and  $V_r e^{j\beta x}$  express a wave propagating toward the direction of



increase (incident wave) and decrease (diffracted wave) of  $x$ , respectively. If we connect a load  $Z_l$  at  $x=\lambda/4$ , from Equation 8.A.1, we get the following relation:

$$\left. \frac{V}{I} \right|_{x=\lambda/4} = Z_0 \frac{1-\Gamma}{1+\Gamma} = Z_l \quad \left( \Gamma = \frac{V_r}{V_i} : \text{voltage reflection constant} \right) \quad (8.A.2)$$

The impedance looking into the transmission line at  $x=0$  becomes

$$Z_i = \left. \frac{V}{I} \right|_{x=0} = Z_0 \frac{1+\Gamma}{1-\Gamma} \quad (8.A.3)$$

From Equations 8.A.2 and 8.A.3, we get the relation  $Z_i = Z_0^2/Z_l$ : a quarter-wavelength line is used as an impedance inverter (converter).

## REFERENCES

- [1] Special issue on software radio. IEEE Commun Mag 1995;33(5).
- [2] Saitoh S, Ebine Y, Tamori T, Kitagawa M, Sintaku M. Basestation equipment; new technical report on digital mobile communication system. NTT DoCoMo Tech J 1993;1:33–38.
- [3] Murota K, Kobayashi K, Nagata K, Chiba K. Mobile station equipment. NTT DoCoMo Tech J 1993;1:43–46.
- [4] Tokuhiko N, Murota K, Sasaki A, Kobayashi K, Nagata K, Chiba K. Portable telephone for personal digital cellular system. Proceedings of the 43th Vehicular Technology Conference; 1993. p 718–721.
- [5] Tarusawa Y, Nojima T. Frequency synthesizer; new technical report on fundamental technologies on mobile communications. NTT DoCoMo Tech J 1994;1:31–36.
- [6] Saito S, Tarusawa Y, Suzuki H. State-preserving intermittently locked loop (SPILL) frequency synthesizer for portable radio. IEEE Trans Microw Theory Tech 1989;MTT-37: 1898–1903.
- [7] Dekker CB. The application of tamed frequency modulation to digital transmission via radio. Proceedings of the IEEE National Telecommunication Conference; 1979. p 55.3.1–55.3.7.
- [8] Kobayashi K, Matsumoto Y, Sakata T, Seki K, Kato S. High-speed QPSK/OQPSK burst modem VLSIC. Proceedings of the IEEE International Conference on Communications; May 1993. p 1735–1739.
- [9] Gosling W, McGeehan JP, Holland PG. Receivers for the Wolfson single-sideband V.H.F. land mobile radio systems. Electron Eng 1979;49:231–235.
- [10] Kahn LR. Single-sideband transmission by envelope elimination and restoration. Proc IRE 1952;40:803–806.
- [11] Petrovic V. Reduction of spurious emission from radio transmitters by means of modulation feedback. IEEE Conference on Radio Spectrum Conservation Techniques; September 1983. p 44–49.
- [12] Petrovic V, Gosling W. Polar-loop transmitter. Electron Lett 1979;15:286–288.

- [13] Akaiwa Y, Nagata Y. Highly efficient digital mobile radio communications with a linear modulation method. IEEE J Selected Area Commun 1987;SAC-5:890–895.
- [14] Ono S, Kondoh N, Shimazaki Y. Digital cellular system with linear modulation. Proceedings of the IEEE Vehicular, Technology Conference; 1989. p 44–49.
- [15] Tomita H. Polar loop linearizer to  $\pi/4$  shift QPSK. Proceedings of the Autumn National Convention of IEICE, No. B-540; 1989.
- [16] Kosh MJ, Fisher RF. A high efficiency 835 MHz linear power amplifier for digital cellular telephony. Proceedings of the IEEE Vehicular, Technology Conference; 1989. p 17–18.
- [17] Uebayashi S, Ohno K, Nojima T, Murata M, Yamada Y. Base station equipment technologies for digital cellular systems. NTT Rev 1992;4:55–63.
- [18] Cox DC. Linear amplification with nonlinear components. IEEE Trans Commun 1974;COM-22:1942–1945.
- [19] Nagata Y. Linear amplification technique for digital mobile communications. Proceedings of the IEEE Vehicular, Technology Conference; 1989. p 159–164.
- [20] Cavers JK. A linearizing predistorter with fast adaptation. Proceedings of the IEEE Vehicular Technology Conference; May 1990. p 41–47.
- [21] Akasaki T, Iwata M, Akaiwa Y. A mathematical expression of nonlinear distortion in RF power amplifier. Proceedings of the IEEE 64th Vehicular Technology Conference, VTC2004-Fall; 2004. p 4217–4220.
- [22] Ding L, Zhou GT. Effects of even-order nonlinear terms on predistortion. Proceedings of the 10th IEEE DSP Workshop; October 2002. p 1–6.
- [23] Vuolevi J, Rahkonen T. *Distortion in RF Power Amplifiers*. Norwood: Artech House; 2003.
- [24] Ghannouchi F, Hammi O. Behavioral modeling and predistortion. IEEE Microw Mag 2009;10:52–64.
- [25] Kim J, Konstantinou K. Digital predistortion of wideband signals based on power amplifier model with memory. Electron Lett 2001;37 (23):1417–1418.
- [26] Morgan D, Ma Z, Kim J, Zierdt M, Pastalan J. A generalized memory polynomial model for digital predistortion of RF power amplifiers. IEEE Trans Signal Process 2006;54:3852–3960.
- [27] Antonio F et al. A novel adaptive predistortion techniques for power amplifiers. Proceedings of the IEEE Vehicular Technology Conference; May 1999. p 1505–1509.
- [28] Ding L et al. A robust digital baseband predistorter constructed using memory polynomials. IEEE Trans Commun 2004;52 (1):159–165.
- [29] Koike S. A set of orthogonal polynomials for use in approximation of nonlinearities in digital QAM systems. IEICE Trans Fundam 2003;E86-A (3):661–666.
- [30] Han S, Lee J. An overview of peak-to-average power ratio reduction techniques for multicarrier transmission. IEEE Wireless Commun 2005;12:56–65.
- [31] Takada T, Muta O, Akaiwa Y. Peak power suppression with parity carrier for multi-carrier transmission. Proceedings of the IEEE Vehicular Technology Conference; 1999. p 2903–2907.
- [32] Doherty W. A new high efficiency power amplifier for modulated waves. Proc IRE 1936;24 (9):1163–1182.
- [33] Akaiwa Y, Koga H. Automatic power control for mobile communication channel. Proc Int Symp Inf Theory Appl 1994;1:487–491.

- [34] Zander J. Distributed cochannel interference control in cellular radio systems. *IEEE Trans Vehicular Technol* 1992;41:305–311.
- [35] Almgren M, Andersson H, Wallstedt K. Power control in a cellular system. *Proceedings of the IEEE Vehicular Technology Conference*; June 1994. p 833–837.
- [36] Okanoué K, Ushirokawa A, Tomita H, Furuya Y. New MLSE receiver free from sample timing and input level controls. *Proceedings of the Vehicular Technology Conference*; May 1993. p 408–411, also in A fast tracking adaptive MLSE for TOMA digital cellular systems. *IEICE Trans Commun* 1994;E77-B:557–565.
- [37] Murota K, Hirade K. GMSK modulation for digital mobile radio telephony. *IEEE Trans Commun* 1981;COM-29:1044–1050.
- [38] Saito S, Suzuki H. Fast carrier-tracking coherent detection with dual-mode carrier recovery circuit for digital land mobile radio transmission. *IEEE J Selected Areas Commun* 1989;7:130–139.
- [39] LaRosa CP, Carney MJ. A fully digital hardware detector for  $\pi/4$  QPSK. *Proceedings of the Vehicular Technology Conference*; May 1992. p 293–297.
- [40] Ikura M, Ohno K, Adachi F. Baseband processing frequency-drift-compensation for QDPSK signal transmission. *Electron Lett* 1991;27:1521–1523.
- [41] Vance IAW. Fully integrated radio paging receiver. *IEEE Proc* 1982;129 (Part F):2–6.
- [42] Vance IAW. Radio receiver for FSK signals. UK Patent Application GB 2057820 A. Filed September 4, 1979.
- [43] Akaiwa Y. Demodulator for digital FM signals. US Patent 4,651,107; 1987.
- [44] Hasegawa M, Mimura M, Takahashi K, Makimoto M. A direct conversion receiver employing a frequency multiplied digital phase-shifting demodulator. *Trans IEICE* 1993;J76-C-I:462–469.
- [45] Kage K, Sasaki Y, Ichihara M, Sato T. The feasibility study of the Nyquist baseband filtered 4-level FM for digital mobile communications. *Proceedings of the Vehicular Technology Conference*; May 1985. p 200–204.
- [46] Nakamura Y, Saito Y. Discriminator with partial response detection of NRZ-FSK signals. *Trans IEICE* 1984;J67-B:607–614.
- [47] Akaiwa Y, Konishi T. An application of the Viterbi decoding to differential detection of frequency-discriminator demodulated FSK signal. *Proceedings of the 4th International Symposium on Personal, Indoor and Mobile Radio Communications*; September 1993. p 210–213.
- [48] Sampei S, Feher K. Adaptive dc-offset compensation algorithm for burst mode operated direct conversion receivers. *Proceedings of the Vehicular Technology Conference*; May 1992. p 93–96.

This page intentionally left blank

## DIGITAL MOBILE RADIO COMMUNICATION SYSTEMS

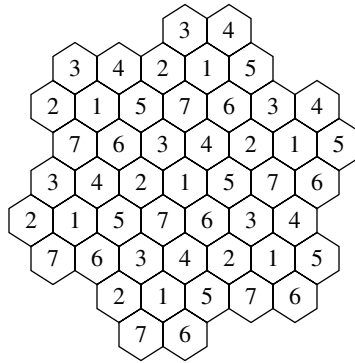
This chapter describes digital mobile radio communication systems. Some basic concepts are discussed first, followed by system descriptions. At the time of writing the first edition of this book, the situation on digital mobile system had been changing rapidly and new systems were appearing. However, in my opinion, the new system evolution rush has passed its peak now. For further description of mobile radio communication systems, the reader may refer to references [1–11] for rather old systems and [12] and [13] for the latest ones.

### 9.1 FUNDAMENTAL CONCEPTS

This section introduces some basic concepts of digital mobile communication systems: cellular systems, multiple access, channel allocation, multiplexing, intercell interference suppression and repeating system are described.

#### 9.1.1 The Cellular Concept

In a cellular system, a service area is covered with many small zones or cells (Fig. 9.1). A zone is illuminated by a base station. With this system, the system capacity, that is, the number of subscribers that can be accommodated, can be increased through reuse of radio channels at different cells where cochannel interference can be neglected. Thus, a cell group is repeated to cover the total



**FIGURE 9.1** A zone layout for cellular systems.

service area. In addition, a lower transmit power is possible because of the smaller coverage of a base station compared to a single-cell system with a large coverage area. On the other hand, the cellular system needs sophisticated procedures for control of location registration for roaming subscribers and call hand-offs between neighboring zones for continuation of a conversation by subscribers moving over zone boundaries. These functions of a cellular system became possible with modern network technologies that include electronic switching and improved radio communication techniques such as frequency-synthesizing transceivers and digital communication.

The technologies supporting a cellular system are classified into two groups: One is related to system control, such as location registration, paging, and hand-offs. The other is related to spectrum efficiency, which is discussed in the following. The spectrum efficiency of a cellular system is given as [14]

$$\eta = \frac{1}{s} \frac{a_c}{2\Delta w n_g n_z} \quad (9.1)$$

where  $a_c$  is the traffic carried per base station,  $s$  is the area size for each cell,  $2\Delta w$  is the spectrum bandwidth per pair of channels,  $n_g$  is the number of cells in one group or the cell cluster size, and  $n_z$  is the number of channels per cell. The system bandwidth becomes  $2\Delta w n_g n_z$ . A smaller cell size can accommodate a higher density of subscribers, and hence, the total capacity of the system increases; however, a larger number of base stations must be deployed. A smaller cell group means a shorter reuse distance. The reuse distance is affected by how rapidly the radio waves attenuate with distance and magnitude of the required protection ratio for cochannel interference.

The term  $a_c/n_z$  reflects how much traffic is carried for a given number of channels. When we assume the Erlang B formula [1],  $a_c$  is given as a function of the number of channels and the blocking probability. A larger number of channels per bandwidth is obtained if we use a narrowband signal. However, a narrowband signal generally requires a higher protection ratio, resulting in a larger  $n_g$ . Thus, finding the optimum bandwidth per channel is not easy.

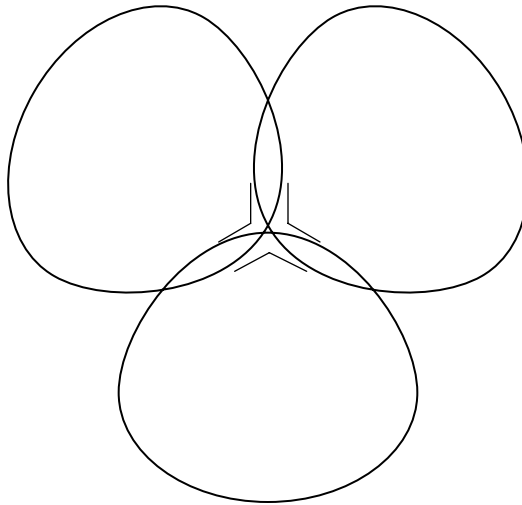
When we assume a hexagonal cell layout, the number of cells in one group is given as [1]

$$n_g = \frac{1}{3} \left( \frac{D}{R} \right)^2 = i^2 + j^2 + i \cdot j \quad (i, j = 1, 2, 3, \dots)$$

where  $R$  is the cell radius and  $D$  is the distance from the center of a cell to the center of the nearest cell reusing the same channel. Determining the cell group number is a difficult task, since it depends on the wave propagation characteristics, including shadowing, zone shape, correlation between desired and interfering signals, and protection ratio [15].

**9.1.1.1 Omni Cell** A cell or zone is the area covered by a base station antenna. When an omnidirectional antenna is used, a zone is covered circularly around the base station. When we plan to place base stations to cover a whole service area by repeating the cell cluster, strategies such as triangular, rectangular, and hexagonal layouts are used. The hexagonal layout is optimal in the sense that the cell cluster size is minimum under the condition of given cochannel interference.

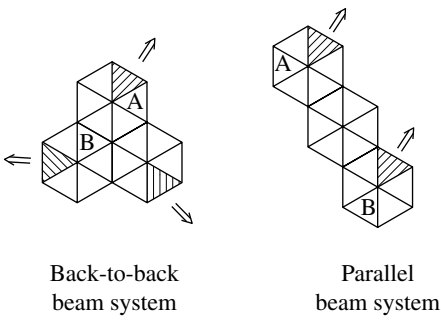
**9.1.1.2 Sector Cell** If a base station employs  $N$  antennas, each of which covers a different direction in the horizontal space,  $N$ -sector cells are produced around the base station, as shown schematically in Figure 9.2. Different channels are used between the sectorized cells. The sector cell layout has advantages in view of efficiency of deploying base stations. An omni cell is split into  $N$ -sector cells, each having a cell size that is  $1/N$  of the omni cell. Thus, a smaller cell is implemented with little effort, that is, the number of base station sites is not increased. Indeed, this technique was introduced to cope with traffic increase in urban areas.



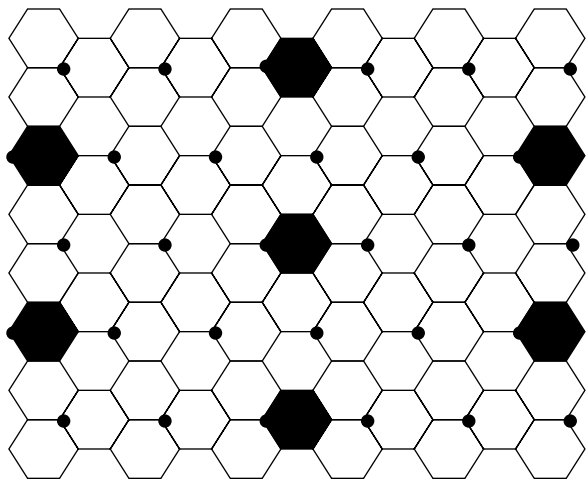
**FIGURE 9.2** Three-sectored cell.

The sector cell layout is classified into two systems, back-to-back and parallel beam systems, as shown in Figure 9.3, where hatched cells use the same channels. Cochannel interference is different in different directions due to the directivity of the sector antenna. Using this fact, an efficient parallel beam sector cell layout was proposed [16] for which the reuse distance differed in the horizontal and vertical directions. An example of a five-site reuse system is shown in Figure 9.4. In this figure, the blacked out sectors, which use the same channels, can be placed closer on the vertical axis than on the horizontal axis, assuming horizontal directional beam antennas. A three-sectored five-site layout is claimed under the condition that the signal to interference power ratio (CIR) is 13 dB.

**9.1.1.3 Beam Tilting** A base station antenna consists of many vertically placed radiating elements, showing directivity in the vertical plane. Beam tilting is a technique in which the antenna directivity is tilted downward to reduce interference



**FIGURE 9.3** Different layouts for six-sectored cell system.



**FIGURE 9.4** Efficient five-site three-sectored cell layout.



to other cells [17]. This technique is also effective in reducing frequency-selective fading by decreasing the multipath echo signal level [18].

**9.1.1.4 Reuse Portioning** In a cell, the signal-to-interference ratio (CIR) is dependent on the location of the mobile station (MS): if the mobile is close to the base station, the CIR is high. This leads to the reuse partitioning concept [19]. The reuse factor is reduced, to be, for example, three in the inner part of the cell shown in Figure 9.5.

Let us assume the reuse is partitioned into two groups where the reuse factor is  $N_A$  and  $N_B$ . Consider a total system consisting of  $S$  channels of which a part  $P$  is assigned to the  $N_A$  group. The number of channels assigned to a cell is then

$$C = \left[ \frac{P}{N_A} + \frac{1-P}{N_B} \right] S$$

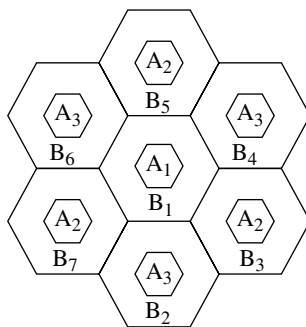
The equivalent reuse factor  $N_{eq}$  is then

$$\begin{aligned} N_{eq} &= \frac{S}{C} \\ &= \frac{N_A N_B}{PN_B + (1-P)N_A} \\ &= fN_A + (1-f)N_B \end{aligned}$$

where

$$f = \frac{PN_B}{PN_B + (1-P)N_A}$$

The equivalent reuse factor changes depending on  $P$  from  $N_A$  ( $P=1$ ) to  $N_B$  ( $P=0$ ).



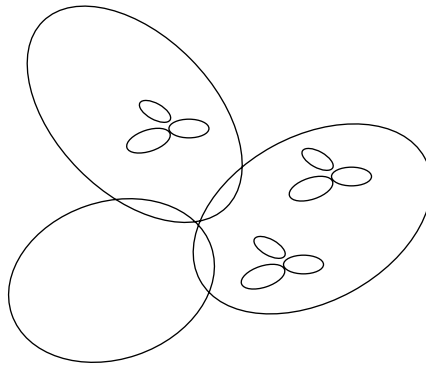
**FIGURE 9.5** Reuse partitioning.

**9.1.1.5 Cell Layout for Nonuniform Traffic Distribution** Traffic density is hardly uniform in all the service areas. It sometimes shows a so-called bell-shaped distribution with its peak in the urban areas. To cope with this, small cell sizes are used in the center areas and cell sizes are gradually increased as the density decreases.

Another case is where traffic density is locally increased in some multiple small areas. In this case, it is not economic to deploy multiple small cells all over the service area. They are used only at the increased traffic areas, and the total area is covered with large (umbrella) cells (Fig. 9.6).

**9.1.1.6 Microcell System** As mentioned earlier, a system with smaller cell size system yields a higher spectrum efficiency (Eq. 9.1). Conventional cells have a radius larger than 1 km. Cellular systems with smaller cell radii, on the order of several hundred meters, are called “microcell” systems; the term “picocell” is even used for cell sizes on the order tens of meters. As the cell size is reduced, the base station antenna height will be lowered to, for example, a lamppost height, and therefore, buildings, roads, and other obstructions will seriously affect the radio wave propagation. The cell shape is then heavily deformed [20] and is barely represented by circular or hexagonal regions. In this situation, it becomes hard to estimate the radio wave propagation and channel reuse planning becomes difficult. In addition, the cells must be highly overlapped to reduce dead-spot areas where service is unavailable because no radio signals reach it. Consequently, the cell cluster size must be increased, resulting in low spectrum efficiency. Furthermore, handovers occur more often for a fast-moving terminal. Solutions to these problems will be discussed in the following.

**9.1.1.7 Microcell and Macrocell Overlaid System** A microcell and macrocell overlaid system (Fig. 9.6) [21, 22] yields some advantages, such as paging and location registration procedures being carried out through macrocell systems, and fast-moving terminals being connected to the macrocells to lessen the hand-off frequency. If the same standards for radio channels (air interface) are used between the microcell and macrocell systems, modifications to mobile terminals can be kept minimal: implementation of automatic power control function is highly desirable to



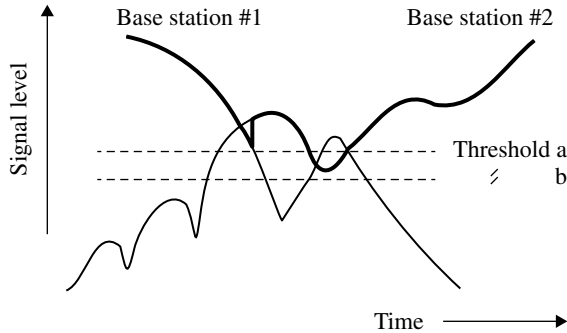
**FIGURE 9.6** A small-cell overlaid with large-cell system.

limit unnecessarily high output power for communications through microcells. Several methods for channel assignment in the microcell and macrocell overlaid system are known: (i) channels are divided between the microcell and macrocell systems; (ii) channels are assigned for a microcell system that are not used in the macrocell system; and (iii) the same channels are used at the same time between the microcell and macrocell systems. The last strategy [23] offers the highest spectrum efficiency. This strategy is possible by increasing the transmit power of the microcell system to overcome the interference from the macrocell system. A rough estimate of the necessary increase in the microcell transmit power is around 5 dB, assuming a cell radius of 3 km for the macrocell and 640 m for the microcell [24]. Since the transmit power of the microcell system is much lower than that for the macrocell system, only interference from the macrocell to microcell is significant and interference in the reverse direction can be neglected.

Deployment of the micro- or picocells under macrocells is discussed under a name of the heterogeneous network in long-term evolution (LTE) systems [12]. Furthermore, much smaller cells called the hemt-cell or home base station are introduced. The home station is intended to compensate for the propagation loss of picocells into buildings. The connection to a backbone network is via a private wire line, which usually inhibits the connection from the third party. The key issue for the heterogeneous network is to enhance a frequency usage efficiency by controlling the signal interference between the different cell systems. To this end, much traffic should be conveyed through the picocells. The picocells suffer signal interference from the macrocells at the cell edge. To suppress the interference, channel allocation schemes under corporation of the macro- and picocells are discussed at the time of writing this book.

**9.1.1.8 Hand-Off** When a mobile terminal in communication crosses the cell boundary and moves into another cell, the call should be connected through the new cell base station. This is called the (intercell) hand-off or handover. The hand-off is initiated when the radio signal quality drops to a certain threshold level. The base station reports the hand-off request to its radio control station, which in turn orders some other base stations to measure and report the signal level of the channel under consideration. The radio channel control station selects the base station that reports the highest received signal level and that has an idle channel(s). If there is no base station that can accommodate the traffic, the hand-off fails and the call is terminated (dropped).

The actual hand-off system is more sophisticated. For example, the hand-off is carried out not with one threshold but with two, to make a hysteresis loop control. To explain this, assume that the signal levels at the boundary region of two coverage areas are as illustrated in Figure 9.7. When the received signal level at base station #1 drops below the threshold level  $a$ , a hand-off to base station #2 is triggered. Although the received signal level at base station #2 drops below the threshold level  $a$ , the hand-off to base station #1 is not triggered as long as the signal level does not drop below the threshold level  $b$  ( $< a$ ). If we do not use this technique, the hand-off is carried out many times during a short time period; this is not desirable in view of the increase in network control procedures and the interruptions to the voice signal.



**FIGURE 9.7** Illustration of hand-off.

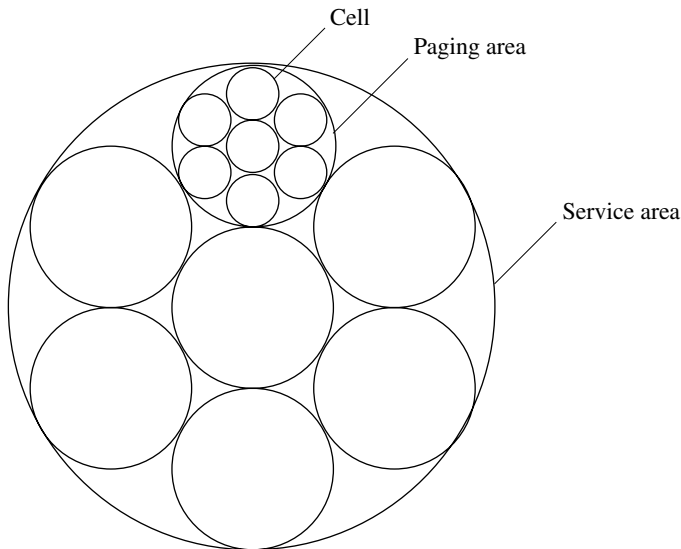
In a time-division multiple access (TDMA) system, an MS takes part in the hand-off procedure by selecting the base station to which the call is handed (the mobile-assisted hand-off). This becomes possible because the MS, in communicating with the current base station, can monitor signals from the candidate base stations for hand-off at the TDMA time slots that are not dedicated to the MS. With this decentralized technique, the hand-off procedure load in the control station is decreased.

Since a forced termination due to a hand-off failure is serious for users, it is important to minimize the probability of a hand-off failure. To this end, schemes such as reserving some channels for hand-off and introducing a queuing system are considered.

**9.1.1.9 Location Registration and Paging** An MS may move all around the service area. When a call arrives for an MS, it is not efficient to page the MS from all the base stations, since the paging traffic becomes large. In a system, the service area is divided into multiple paging areas as shown schematically in Figure 9.8. An MS is paged from a paging area. To know which paging area is responsible to an MS, each MS is registered to a paging area. This location registration is carried out when an MS receives a paging area identification code that is different from the one for the last location registration. If the paging area size becomes smaller, the location registration traffic increases. Therefore, the paging area sizes should be decided by compromising the traffic for paging and for location registration.

The location information is stored in a file, the home location register, at a switching center. When an MS moves into the territory of another switching center, the location information of the MS is recorded in a file, the visitor location register, and is also sent to the home switching center to record this fact.

Even in an analog cellular system, a digital data transmission system is necessary for system control such as paging, location registration, and call setup and release. In this system, the data rate is slow because digital data is transmitted through an analog voice channel. This fact places a burden on the system control. A high data rate (HDR) transmission capability of a digital cellular system can relieve the problem.



**FIGURE 9.8** Cell, paging area, and service area.

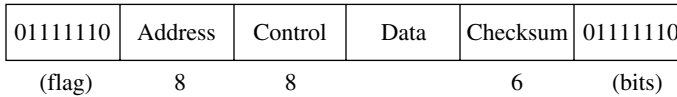
As the paging area size becomes small, the frequency of location registration grows. In addition, location registration is made backward and forward frequently between the boundaries of location registration areas. In order to prevent such a phenomenon, the overlaid location registration area is adopted in the NTT digital cellular system [25]. In this system, boundaries of the location registration areas in a layer are placed on the center of those in the other layers.

A technique was proposed in Ref. [26] for autonomous creation of location registrations and paging areas. A fully distributed location registration strategy is discussed in Ref. [27].

## 9.1.2 Multiple Access

The communication media, the radio channels in mobile communication, are used by multiple users. Multiple-access methods are classified into random or contention access schemes and controlled access schemes. The controlled access schemes include the demand assignment system, the polling system, and the token-passing system. In the demand assignment system, a channel is assigned on demand. This scheme is used for telephone systems. The others are used for data communication systems. In a polling system, a central station polls subscriber stations in turn and only a polled station can get access to the channel. In a token-passing system, a subscriber station that has received a token has the right to access the channel. The token is passed to the next station when a transmission ends.

**9.1.2.1 Random Access System** Random access is used for data communications or in the reservation phase of an on-demand channel assignment system. In a random access system, collision may occur between signals transmitted from different stations.



**FIGURE 9.9** Frame format for a packet.

If the stations in conflict retransmit immediately after a collision, the collision continues forever. To avoid the succession of collisions, the signal is retransmitted after a back-off time, which is randomly selected from a given time interval.

In a random access data transmission system, the data are transmitted in a form of a so-called packet, which is a block of data consisting of a flag pattern and fields for control data, address, information data, and checksum for error detection (Fig. 9.9). In terms of switching technologies, a conventional telephone system is called a circuit-switched system, where the channel is held throughout a conversation; packet data transmission is called a packet-switched system. In a packet transmission system, the average throughput and average transmission delay are important measures of performance. The throughput, denoted by  $S$ , is defined as the rate of effective channel usage per unit time. The delay, denoted by  $D$ , is the average time between generation and reception of a packet. These performance measures are discussed in the following [28].

Let us assume random arrivals of packets. The probability that  $k$  packets arrive during a packet time length  $T$  is given by (Appendix 9.A)

$$P_k(T) = \frac{(\lambda T)^k}{k!} e^{-\lambda T}$$

where  $\lambda$  is packet arrival rate per unit time including packet retransmissions. Let  $G = \lambda T$  denote the average number of packets to be transmitted.

The ALOHA system is a famous random access system. It has two versions: pure ALOHA and slotted ALOHA. In the pure ALOHA system, when a packet is generated, a station sends the packet immediately. In the slotted ALOHA system, packet transmissions are synchronized to time slots, which are known throughout the system.

The throughput  $S$  is given by the product of  $G$  and the probability that a packet experiences no collisions. For the pure ALOHA system, the probability for a packet to meet no collision is given by  $P_0(2T)$ . The throughput  $S(<G)$  is then

$$\begin{aligned} S &= GP_0(2T) \\ &= Ge^{-2G} \end{aligned}$$

The average number of packet transmissions for a packet to be successfully transmitted is  $G/S - 1$ . The average transmission delay is then

$$D = T + 2\tau + (e^{2G} - 1)(T + 2\tau + B),$$

where  $\tau$  is the round-trip time and  $B$  is the average back-off time at collision.

For the slotted ALOHA system, the probability that a packet meets no collision is given by  $P_0(T) = e^{-G}$ , which is smaller than that for the pure ALOHA system for a fixed  $G$  ( $e^{-G} > e^{-2G}$ ). The average throughput in this case is

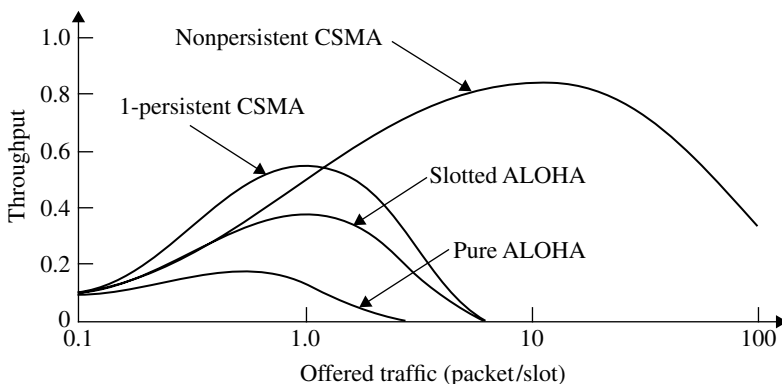
$$S = Ge^{-G}.$$

A packet waits on average  $T/2$  until the next slot starts. Let us assume a collided packet waits for a random number of time slots between 0 and  $K-1$ ; then the average back-off time  $B$  is  $(K-1)T/2$ . Thus, the average transmission delay is

$$D = 1.5T + 2\tau + (e^G - 1)[(K+1)T/2 + 2\tau].$$

**9.1.2.2 Carrier Sense Multiple Access (CSMA) System** In this system, stations measure the carrier level of the channel to monitor whether the channel is busy or not. If the channel is sensed to be idle, the station sends a packet, and otherwise, it waits for another chance. The carrier sense is not complete in detecting whether the channel is busy or not, since there is a delay for signals transmitted from other stations. The CSMA schemes are classified into three subgroups: nonpersistent CSMA, 1-persistent CSMA, and  $p$ -persistent CSMA. The nonpersistent system obeys the back-off algorithm if the channel is sensed busy. In the 1-persistent system, the packet is sent immediately after the time when the channel is idle. In the  $p$ -persistent system, even when the channel is sensed busy, the packet is transmitted with a probability of  $p$  or waits for  $t$  time units with a probability of  $1-p$  and senses the carrier again. The case  $p=1$  is the 1-persistent system.

The CSMA is improved with collision detection (CSMA/CD). In this system, carrier sensing is carried out even during signal transmission. If a collision is detected, transmission is stopped so as not to waste the channel. Carrier sensing performs well in cable transmission systems. However, it may become uncertain in wireless systems, since the radio waves are not guaranteed to reach all the stations involved in the current communication, because of radio wave propagation uncertainty. The throughput versus offered traffic load is shown in Figure 9.10 [29].



**FIGURE 9.10** Throughput versus offered traffic.

### 9.1.3 Channel Assignment

In some mobile telephone systems, a radio channel is assigned for a call on demand. In a cellular system, the channel assignment for calls in each cell is important since it directly affects the efficiency of channel reuse between different cells. The channel assignment schemes are divided into three categories: fixed channel assignment, dynamic channel assignment, and hybrid channel assignment. In the fixed channel assignment, channels are fixed for each cell and the channels are never used by the other cells in the same cell cluster. In the dynamic channel assignment, channels are dynamically used between cells depending on the situation. In the hybrid channel assignment, some channels are fixed for each cell and others are used dynamically.

The fixed channel assignment scheme requires channel assignment planning in advance according to cell layout design. The cell layout design and the channel assignment are probably the most important technical matters in implementing an actual cellular system. This is so because radio propagation is in practice not uniform in contrast to the idealized situation that we often assume in theoretical analyses.

The dynamic channel assignment systems were known long ago. The first available ones, the mean square, the nearest neighbor, and the channel borrowing methods appeared in reference [1]. In recent years, the dynamic channel assignment schemes have received attention as microcell systems have become a hot topic: their aspect of adaptive or decentralized operation is recognized as necessary to channel assignment in microcell systems, where radio wave propagation is too irregular to allow estimation of interference between cells. Under this condition, we should not design the cell layout and the channel assignment in advance.

**9.1.3.1 First Available** In the first available or the random assignment scheme, when a call arrives at a cell, the cell base station searches for an available channel in a given order or at random. The examination of the channel is simply done by monitoring the received signal levels: if the level is lower than a threshold level, the channel can be used. Thus, channels are selected from all the channels at each cell with its own decision depending on the channel usage status in the interfering cells. Channels are assigned autonomously in a decentralized manner. Consequently, the term “channel assignment” would be better replaced with the concept of channel “selection.” Thus, we are free from channel assignment planning. Not all of the dynamic channel assignments are autonomous or decentralized. The performance of the first available system is very different with the different schemes for searching for a channel, namely, random search and ordered search. The random search system has a lower channel reuse performance than the ordered search system since no systematic channel reuse is obtained. A version of the first available system with ordered search is known as the autonomous reuse partitioning (ARP) system and will be discussed later.

**9.1.3.2 Channel Segregation** The channel segregation [30, 31] was the first method where the learning concept was introduced to the channel assignment system. In this system, each cell learns, through interference and channel usage between



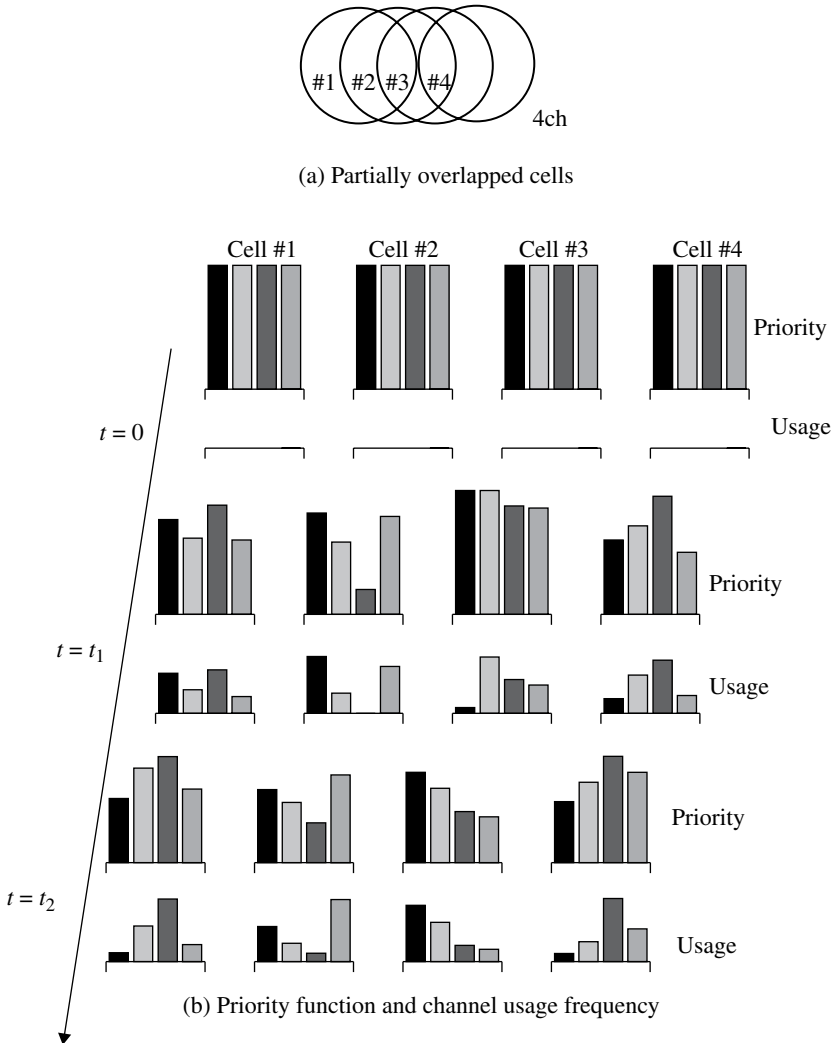
other cells, to acquire some favorite channels. Channel reuse is established autonomously. The term “channel segregation” comes from the analogy with ecological territories of animals. The territories (channel usage in each cell) are established through several contests (interference) between the owners (cells). This, termed “segregation,” can be understood as a system where resources are shared in a decentralized manner by many users. The channel segregation system is described as follows. Each base station has a table in which a priority function of channels is stored; a higher priority is given to a channel with a higher value of the priority function.

1. When a call arrives at a base station, the channel with the highest priority is chosen from the channels that are not used in the base station.
2. The channel is tested for whether it can be used for the new call. Specifically, the signal level of that channel is measured at the base station and/or at the MS. If the measured signal level is below a given threshold level, the channel is decided idle; otherwise, it is decided busy.
3. If the channel is decided idle, the priority of the channel is increased and the channel is used.
4. If the channel is decided busy, its priority function is decreased and the next highest priority channel is set to be tested.
5. Steps 1 to 4 are repeated. If there is no available channel, the call is blocked.

The estimated carrier to interference level instead of the received signal level may be a better criterion for the test of a candidate channel. It is important that the aforementioned procedure is made at each base station independently. There is neither information exchange nor central control information between base stations.

The most interesting question is whether the channels segregate automatically between cells. To investigate this computer simulation experiment is done (see Fig. 9.11). Random calls are originated in each cell. As time proceeds, each cell acquires a different channel; in other words, channel segregation is established. We consider a one-dimensional system, where cells are partially overlapped as shown in Figure 9.11a. Cells #1 and #4 never overlap, that is, no interference occurs between them. The learning process in terms of the priority function and channel usage frequency is shown in Figure 9.11b. The interfering cells capture different channels and cells #1 and #4 capture the same channel. Thus, the channel reuse is automatically established. The channel segregation is also confirmed in two-dimensional systems. Owing to the structured channel reuse pattern, channel segregation shows a better blocking probability performance than the random assignment system.

Channel segregation was applied for the TDMA/frequency-division multiple access (FDMA) system in [32]. In a TDMA/FDMA system, time-division as well as frequency-division channels are prepared. If we apply the random channel assignment to a TDMA/FDMA system, the problem occurs that a call can be blocked even when a radio channel is idle. This happens when all channels on a carrier frequency are used and the idle channel is on another carrier frequency; this occurs because of the inability



**FIGURE 9.11** Channel segregation.

of a base station transceiver to switch its carrier frequency time slot by time slot so that it therefore cannot access the idle channel. The probability of the occurrence of this inaccessible channel problem is high when time- and frequency-division channels are used at random. The computer simulation results of a TDMA/FDMA system with channel segregation show that the call blocking probability due to the inaccessible channel problem is decreased. This is the result of the bunching effect, where channels to be used in a base station tend to gather on the same frequency.

The microcell and macrocell overlaid system is another example where channel segregation has been applied [24]. The microcell systems use dynamically the same channels that are assigned to the macrocell system, even when those channels are

being used in the macrocell system as described in Section 9.1.1. On learning the cochannel interference from the macrocell system, the microcell system tends to use the channels that show the least interference from the macrocells.

When multiple transmitters/receivers are located close to each other, a problem called intermodulation occurs: interfering signals are produced due to nonlinearity of the transceiver circuits. The frequencies of these interfering signals are given by the algebraic sum of the carrier frequencies and their harmonics. We must carefully select a set of carrier frequencies that will not be interfered with by the intermodulated signals. Channel segregation can successfully avoid the intermodulation interference by automatically selecting an appropriate set of carrier frequencies [33].

**9.1.3.3 ARP** The ARP (autonomous reuse partitioning) dynamic channel assignment method [34] automatically establishes the reuse partitioning (Section 9.1.1). The steps in this method are simple:

1. Channels are ordered identically throughout all base stations.
2. When a call arrives at a base station, the channels are tested for whether they can be used or not in that order. If a channel shows carrier-to-interference ratio (CIR) above a given threshold value in both the forward and backward link, the channel is determined idle. The first available channel is used.
3. If there is no available channel, the call is blocked.

This is a fully decentralized system. The evidence for the establishment of automatic reuse portioning is shown in Figure 9.12. In this figure, computer simulation result of the average number of simultaneous calls versus channel number is shown. We can see that the lower-numbered channels are more densely used throughout the service area: this means that the reuse distance for the channel is shorter than that for a channel with a higher number channel. Even though the reuse distance is short for a low-numbered channel, the CIR value is similar to that for high number channel as shown in Figure 9.13.

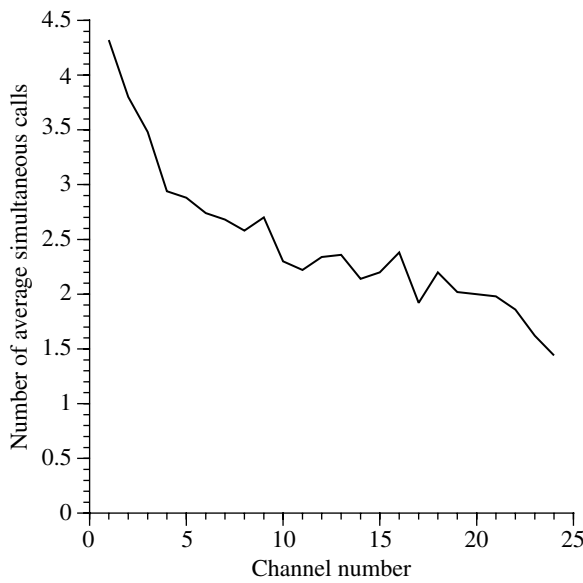
Self-organized reuse partitioning (SORP) [35] was proposed for decreasing the channel sensing times in ARP. In this method, the average power for each channel is calculated at each base station using the signal levels at the base stations. The channels are sensed from those showing an average power level that is closest to the mobile transmit power.

Another method for decreasing the number of channels sensed with ARP was proposed in Ref. [36], where the starting channel for carrier sensing is adaptively selected depending on the received signal level at the base station. At each base station, the average received signal level is calculated and stored for each channel: consider channel  $j$ , then the average received signal level  $S(j)$  is calculated as

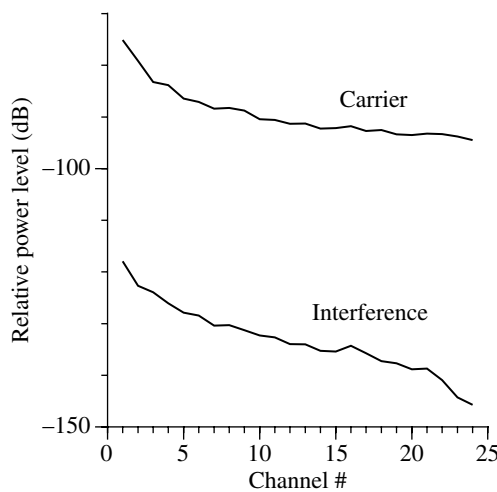
$$S(j) = \frac{nS(j) + P}{n + 1}$$

where  $n$  is the number allocation of that channel so far and  $P$  is the received signal in dB at the current location. The starting channel is the channel that has the average received signal level larger than the current signal level by a given margin.

SORP is combined with channel segregation in [37].



**FIGURE 9.12** Average number of simultaneous calls versus channel number with autonomous reuse partitioning method.



**FIGURE 9.13** Average carrier and interference signal level versus channel number with autonomous reuse partitioning method.

**9.1.3.4 Packet Reservation Multiple Access (PRMA)** This method [38] is closely related to the reservation ALOHA method. The channel is divided into some number of time slots to form a frame. Voice signals are transmitted only at talk spurts in the form of a packet, so a voice activity detector is employed at the voice terminal. In order to avoid collision between a voice packet and other packets, a time slot is

reserved by a voice terminal for each talk spurt period. At the beginning of a talk spurt, the voice terminal makes a reservation for a time slot through a random access slot that is not reserved. The status “reserved” or “available” is broadcast from the base station. The speech signal waits until the reservation becomes successful. If the reservation is not successful within 32 ms, then the terminal discards the speech packet. This distinguishes PRMA from reservation ALOHA.

Since speech signals have a talk spurt rate of around 40%, the radio channel can be used effectively by statistically multiplexing speech signals in the PRMA system: the number of served radio channels can be less than the number of speech terminals in conversation. The statistical multiplexing technique in a TDMA system is known as digital speech interpolation (DSI) [39], which requires complex control; on the other hand, the DSI effect is easily embedded in the PRMA system. In applying the PRMA method to digital cellular systems, a method combining the PRMA and channel segregation was proposed in Ref. [40].

For other methods, refer to Refs. [41–45].

**9.1.3.5 Channel Allocations in Wireless Packet Data Transmission Systems** In a wireless communication system, channel qualities differ for users due to propagation loss and fading. Furthermore, the channel qualities for a user changes in time due to the fading and signal interference from other users. In a voice communication system, useful channel allocation methods are restricted because of the requirement for the real-time communication. In contrast to this, various channel allocation methods are applicable to a data transmission system, since significant time delay is allowed in data packet transmission. Complying a high overall performance of a wireless packet system and fairness in the channel allocation for users with different channel qualities is an important issue in a wireless packet system.

The importance of channel allocation scheduling was pointed out, and an efficient scheme was proposed first in a wireless packet system for HDR system by Qualcomm Inc. [46, 47]. The method is called proportional fair sharing (PFS), where the channel is allocated to a user who has the highest priority based on a criterion that is defined for each user as a ratio of the transmission rate attainable at a current channel quality to the so far averaged transmission rate. It complies with the requirements for the overall transmission rate and the channel allocation fairness in a system. This method [47], [48] is related to the proportional fairness (PF) criterion given in Ref. [49]. It is shown that the PFS fulfills the PF criterion if transmission speeds averaged under fading are considered [50].

Here, the channel allocation scheme based on the PF criterion is described first. Then the other methods including round-robin (RR), the maximal throughput, and their modification are discussed. The throughput performance is given under the Rayleigh fading with a simple assumption that the transmission rate is proportional to the signal-to-noise power ratio (S/N) in the channel.

Denoting the S/N by  $q_i$  for user  $i$ , the transmission rate  $x_i$  is given by the assumption as

$$x_i = kq_i \quad (i = 1, 2, \dots, N)$$

where  $k$  is a constant and  $N$  is the total number of users. Denoting channel allocation rate for user  $i$  by  $p_i$ , the average transmission rate becomes

$$y_i = \langle x_i \rangle = p_i x_i \quad \left( \sum_{i=1}^N p_i = 1 \right)$$

where we assume  $x_i$  and  $p_i$  never change in a time period for the time averaging. We assume that each user's data are always in a waiting mode for transmission (full buffer).

**The PF Criterion and Its Algorithm** The channel allocation scheme to fulfill the PF criterion is defined such as that the summation of relative change from the optimum ones  $y_i^*$  for other allocations  $y_i^*$  becomes negative or zero [49], that is,

$$\sum_{i=1}^N \frac{y_i^* - y_i}{y_i} = \sum_{i=1}^N \frac{p_i^* - p_i}{p_i} \leq 0 \quad \left( \sum_{i=1}^N p_i = \sum_{i=1}^N p_i^* = 1 \right)$$

where  $y_i^* = p_i^* x_i$ . We wish to get  $\{p_i\} = (i=1, 2, \dots, N)$  to achieve such channel allocation. Letting

$$\Delta p_i = p_i^* - p_i \quad \left( \sum_{i=1}^N \Delta p_i = 0 \right)$$

we should find  $\{p_i\}$  to satisfy  $\sum_{i=1}^N (\Delta p_i / p_i) = 0$ . We follow the discussion in Ref. [49]. We obtain  $\{p_i\}$  to maximize the target function

$$U(p_1, p_2, \dots, p_N) = \sum_{i=1}^N \log p_i x_i$$

subjected to  $\sum_{i=1}^N p_i = 1$ .

We use Lagrange multiplier method as

$$L(p_1, p_2, \dots, p_N) = \sum_{i=1}^N p_i x_i - \lambda \left( \sum_{i=1}^N p_i - 1 \right)$$

where  $\lambda$  is an indeterminate coefficient. Letting the partial differentiation of the above equation to be zero, we have

$$\frac{\partial L}{\partial p_i} = \frac{x_i}{p_i} - \lambda = \frac{1}{p_i} - \lambda = 0 \quad (i=1, 2, \dots, N)$$

Thus, we obtain  $p_i = 1/\lambda$  ( $i=1, 2, \dots, N$ ). From  $\sum_{i=1}^N p_i = 1$ , we have  $\lambda = N$ .

### ***The Other Allocation Methods***

**RR** Channel is allocated to a user in turn, and therefore  $p_i = 1/N$  ( $i = 1, 2, \dots, N$ ). The result is the same as the one with the PF as shown later.

**The Maximum Throughput** Channel is allocated so as to maximize the overall throughput. Therefore, we have  $p_m = 1$  and  $p_{i \neq m} = 0$ , where the user  $m$  is such that  $x_m = \max\{x_1, x_2, \dots, x_N\}$ .

**PFS** Channel is allocated to a user  $m$ , who has the maximum value of  $\theta_m$  that is defined for each user as the ratio of a transmission rate given with current channel quality to the average channel rates so far, that is,  $p_m = 1$  and  $p_{i \neq m} = 0$  where  $\theta_m = \max\{\theta_1, \theta_2, \dots, \theta_N\}$ .

**Transmission Rates** We consider two users in the Rayleigh fading channel for simplicity. The received signal powers  $\gamma_1$  and  $\gamma_2$  for users 1 and 2 are, respectively, given as (Section 4.3)

$$p(\gamma_1) = \frac{1}{b_1} e^{-\gamma_1/b_1},$$

$$p(\gamma_2) = \frac{1}{b_2} e^{-\gamma_2/b_2}$$

where  $b_1$  and  $b_2$  are average received power, that is,  $b_1 = \langle \gamma_1 \rangle$  and  $b_2 = \langle \gamma_2 \rangle$ .

We assume  $b_1 \geq b_2$  for simplicity.

Assuming the equal noise power for the users and denoting it by  $N$ , the signal noise power ratio becomes  $q_i = \gamma_i / N$  ( $i = 1, 2$ ). The transmission rates are given as  $x_i = kq_i = k\gamma_i / N = k'\gamma_i$  ( $k' = k / N$ ).

The average transmission rates are calculated in the following.

**RR**

$$\begin{aligned} \langle y_i \rangle &= \frac{k'}{2} \int_0^\infty \gamma_i p(\gamma_i) d\gamma_i \\ &= \frac{k'}{2} \langle y_i \rangle = \frac{k'}{2} b_i \quad (i = 1, 2) \end{aligned}$$

**The maximum throughput** Using the following relation

$$p_1 = \begin{cases} 1 & (\gamma_1 \geq \gamma_2) \\ 0 & (\text{otherwise}) \end{cases} \quad \text{and} \quad p_2 = \begin{cases} 1 & (\gamma_1 < \gamma_2) \\ 0 & (\text{otherwise}) \end{cases}$$

we have

$$\begin{aligned}\langle y_1 \rangle &= k' \int_0^\infty d\gamma_1 \gamma_1 p(\gamma_1) \int_0^{\gamma_1} p(\gamma_2) d\gamma_2, \\ \langle y_2 \rangle &= k' \int_0^\infty d\gamma_2 \gamma_2 p(\gamma_2) \int_0^{\gamma_2} p(\gamma_1) d\gamma_1.\end{aligned}$$

Performing the integrals, we gets

$$\begin{aligned}\langle y_1 \rangle &= k'b_1 \left[ 1 - \frac{1}{(1 + 1/\beta)^2} \right], \\ \langle y_2 \rangle &= k'b_2 \left[ 1 - \frac{1}{(1 + \beta)^2} \right]\end{aligned}$$

where  $\beta = b_2 / b_1$ .

With this allocation, a user with a quite low received power scarcely gets a chance for transmission. To give a better chance for the user, we modify the algorithm as

$$p_1 = \begin{cases} 1 & (\gamma_1 / b_1 \geq \gamma_2 / b_2) \\ 0 & \text{(otherwise)} \end{cases} \quad \text{and} \quad p_2 = \begin{cases} 1 & (\gamma_1 / b_1 < \gamma_2 / b_2) \\ 0 & \text{(otherwise)} \end{cases}$$

then we have

$$\langle y_1 \rangle = \frac{3}{4} k'b_1 \quad \text{and} \quad \langle y_2 \rangle = \frac{3}{4} k'b_2.$$

*PF* The results become the same as for the RR method. We modify this algorithm to give a high priority for a user with a low received power as  $\frac{p_1}{p_2} = \frac{b_2}{b_1} (= \beta)$ , and

then we have  $p_1 = \frac{\beta}{1 + \beta}$  and  $p_2 = \frac{1}{1 + \beta}$

The average transmission rates are given as

$$\begin{aligned}\langle y_1 \rangle &= \frac{\beta}{1 + \beta} k'b_1, \\ \langle y_2 \rangle &= \frac{1}{1 + \beta} k'b_2 = \frac{\beta}{1 + \beta} k'b_1 (= \langle y_1 \rangle).\end{aligned}$$

*PFS* Solving difference equations, the convergent solutions are given as [48]

$$\langle y_1 \rangle = \frac{3}{4} k'b_1 \quad \text{and} \quad \langle y_2 \rangle = \frac{3}{4} k'b_2$$



This results are the same as those with the modified maximum throughput method discussed before.

To compare performance of the channel allocation methods, we use the total throughput  $\langle y_t \rangle = \langle y_1 \rangle + \langle y_2 \rangle$  and define a parameter that expresses the degree of fairness as  $d_f = \langle y_1 \rangle / \langle y_2 \rangle$ . If we normalize  $\langle y_1 \rangle$  and  $\langle y_2 \rangle$  with  $k'b_1$  or letting  $k'b_1 = 1$ ,  $\langle y_t \rangle$  and  $d_f$  are expressed only with  $\beta$  as shown in Table 9.1.

Figure 9.14 shows the total throughput and the degree of fairness as a function of  $\beta$ . The maximum throughput MT shows the highest total throughput; however, its fairness is worst. The modified MT (MT') and the proportional fair scheduling (PFS) show the second high total throughputs and a moderate fairness. Although their fairness performance is the same as that for RR and PF, they give a 1.5 times higher total throughput than RR and PF. This fact is owing to the multiuser diversity effect (Section 7.1) [51, 52], which is obtained through dynamic channel assignment for users depending on their channel qualities. The PF allocation as well as RR in principle never takes into consideration the comparison results of instantaneous channel qualities between users. The modified PF (PF') shows equal fairness irrespective to the average power ratio, sacrificing the total throughput.

### 9.1.4 Multiple-Access System

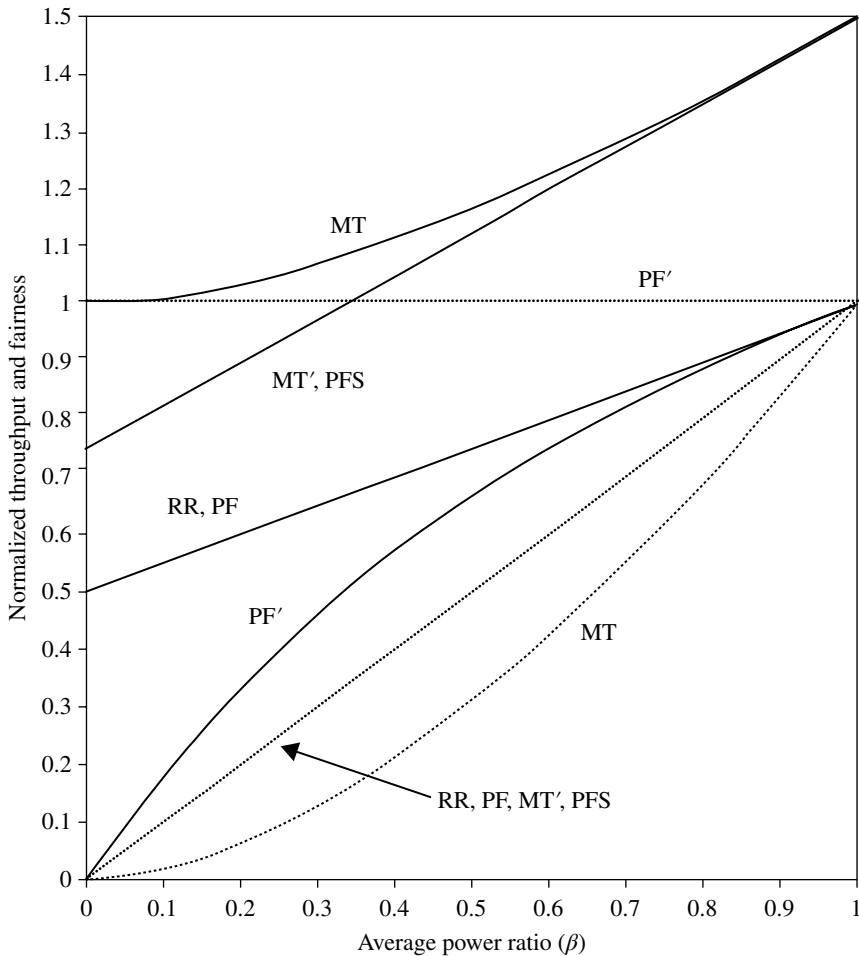
Many channels must be prepared to accommodating a number of users in a circuit-switched system. We have three ways to produce channels using a given bandwidth: frequency –, time –, and code – division multiplexing. When we consider the uplink (subscriber to base station), the prepared channels are accessed by many users (multiple access). Thus, we have FDMA, TDMA, and CDMA systems. The radio transmission system differs considerably between these multiple-access systems, so that which multiple-access system is selected is the subject of heated debate.

**TABLE 9.1 Performance comparison**

	$\langle y_t \rangle$	$d_f$
RR	$\frac{1}{2}(1 + \beta)$	$\beta$
MT	$\frac{1 + \beta + \beta^2}{1 + \beta}$	$\frac{(2 + \beta)\beta^2}{1 + 2\beta}$
MT'	$\frac{3}{4}(1 + \beta)$	$\beta$
PF'	$\frac{2\beta}{1 + \beta}$	1

The proportional fairness (PF) and the proportional fair sharing (PFS) show the same results as for RR and MT', respectively.

RR, round-robin; MT, maximum throughput; MT', modified MT; PF', modified proportional fairness,  $\beta = b_2/b_1$ .



**FIGURE 9.14** Normalized throughput and fairness (solid line, throughput; broken line, fairness).

FDMA is used for analog mobile radio systems. This system is implemented with minimum signal processing. The number of channels given with this system is decided by the carrier frequency spacing. Since the channels are split (orthogonally) into frequency divisions, a stable carrier frequency source and a channel selection filter with a sharp roll-off transfer function are required. A narrower spectrum bandwidth of the modulated carrier signal is advantageous. The modulated signal bandwidth per user for digital cellular systems becomes very narrow owing to the evolution of low-bit-rate voice coding and spectrum-efficient linear modulation. Furthermore, a higher frequency band tends to be used. In this situation, a guard band between channels given for discrimination in frequency division becomes relatively crucial so as not to decrease the spectrum efficiency. Although FDMA systems have the advantage that

they are immune against frequency-selective fading due the narrow spectrum, they have not been favored in recent arguments about digital mobile communication systems. This is because FDMA systems are not superior to TDMA or CDMA systems in other aspects described later. The OFDMA system (Section 6.5), which uses multiple subcarriers, falls into the OFDM category and is adopted to recent wireless systems such as LTE system combined with dynamic channel allocation.

TDMA systems have become feasible for digital communication systems, where voice signals can be converted into digital signals and therefore the multiplexing in time division is an easy task. Channels are given periodic time slots in a period time called a frame. Since the channels are orthogonal in time, it is easy to select a channel: we merely open a gate at the wanted time slot. Furthermore, no interference between the time-division channels occurs even if they are subjected to a nonlinear channel as long as a single carrier (SC) is assumed. The number of base station transceivers is decreased by a factor of  $1/N$ , where  $N$  is the number of time-division multiplexed channels on a carrier.

Since  $N$  channels are multiplexed, the total transmission data rate becomes  $N$  times the rate of a single channel. The increase of the data rate requires  $N$  times the bandwidth and  $N$  times the peak power as compared to an FDMA system: however, the bandwidth per channel and the average power are the same. A high channel speed may cause intersymbol interference in a frequency-selective fading channel.

In a TDMA system, time-slot synchronization in a frame becomes important since, if the synchronization is lost, the channels may collide with each other. In order to establish the synchronization, the base station periodically sends a frame timing signal which subscriber stations use for synchronization. When subscriber stations are located in a coverage area with different distances from the base station, the signal propagation delay difference becomes a problem. A solution to problem is to adjust the subscriber timings by sending a control signal from the base station: at the initial stage of the synchronization procedure, the subscriber station sends a burst signal that is short enough to avoid collisions even when the subscriber is located at the most distant place in the predetermined coverage area. The base station then measures the timing error for that station and feeds back the information to that station. To avoid the effects of residual errors in this timing control system, guard times are placed between time slots. This guard time and the preamble signal, which bears the fixed signal, are the overheads that reduce the spectrum efficiency of TDMA systems.

CDMA systems use different spectrum spreading codes to discriminate channels from each other. Direct sequence (DS) and frequency hopping (FH) are known as spread-spectrum (SS) systems. Consider first the DS system. If we use orthogonal codes such as Walsh functions (Section 2.1.3),  $N$  channels are orthogonally multiplexed on the same frequency and at the same time using  $N$  times the bandwidth of a channel before spreading. A CDMA system enjoys the benefits of SS communications: a high time resolution and wide bandwidth. Furthermore, no management of frequencies or time slots is required for this system.

Since CDMA systems are based on the orthogonality between spectrum spreading codes, the degree of correlation becomes crucial for actual systems. A residual cross

correlation between codes and timing differences between codes reduce the orthogonality. When applied to mobile radio communications, the near-far problem accelerates the loss of orthogonality. For this system, automatic transmit power control, which keeps the received signal level constant at the base station, is necessary.

FH systems are considerably different from DS systems in terms of the aforementioned argument. For a slow FH system, channel orthogonality is similar to that of an FDMA system, due to the difference in frequencies. FH systems can randomize the cochannel interference and fading effects.

### 9.1.5 Intercell Interference Suppression

A cellular system is never free from intercell (cochannel) interference, since it reuses simultaneously the same channel at different cells. The effect of the interference becomes high for mobile terminals that locate near to cell edges, since the desired signal level drops and the interference signal level becomes high. Some methods are known to suppress the interference with an adaptive antenna array or digital signal processing as described in Section 7.6. An SS system can suppress the interference at the receiver output proportionally to its spreading factor (Section 6.4). The spreading factor should be decreased as the data rate increases to keep a fixed bandwidth, and the interference suppression becomes less. Furthermore, the orthogonality between spreading codes decreases if the channel becomes dispersive. A high orthogonality between signals is easily obtained in a frequency-division or a time-division system. To use effectively the high orthogonality between signals, channel allocation between different cells becomes important so as not to reuse the same channel at a high interference level. To accomplish such a channel allocation, a channel allocation plan should be carried out, or channel usage information between interfering cells should be exchanged.

To enhance channel quality for an MS at a cell edge, some methods are known, where neighboring cell base stations become cooperative to send/receive signal to/from the MS [53]. A method is that neighboring base stations send the same signal to the MS. This is a site transmission diversity system that increases the received power at the MS without causing cochannel interference. The second method is that a base station is selected to send to the MS depending on the channel quality. This is a fast selection site transmit diversity system. The other method is that a base station employs an adaptive antenna system to form a radio wave beam that has a suppressed emission to an interfered MS(s) connected to another cell base station.

For an uplink, multiple base stations receive signals from an MS in a cooperative manner, that is, site reception diversity with selection or combining (Section 7.1).

### 9.1.6 Repeater System

Deployment of a repeater station is effective when the distance between a base station and MSs becomes long and the signal attenuation is high. If we can use new frequencies for the link between the base and repeater stations, we can cope flexibly with any repeating system. However, an extra frequency band is hard to get for an

existing system. Then frequency channels that are prepared for communication between a base station and MSs should be used.

The simplest repeating method is that a received signal is just amplified and forwarded (AF). The defect of the method is its inherent instability due to feedback of transmit signal into the input of the repeater. To assure a stable operation (Section 2.3.2), we should design the repeater system such that the isolation between the output and the input of the repeater is kept sufficiently high, an amplifier gain is decreased, and the feedback signal level is suppressed with an appropriate signal processing. The addition of noise at the repeater is the other defect with the AF method.

Other repeating methods to avoid the inherent stability problem are such that different frequencies are used for receiving and transmitting, or a channel is used in a time-division duplexing mode. The disadvantage of the methods is their low efficiency (one half of that for AF method) in frequency use, since, for a single signal transmission, one half of the frequency channels and one half of a time-division channel are used in the former method and the latter one, respectively.

The noise power increase in a repeater with such as an AF system can be avoided by a decode-and-forward (DF) system, where a received signal is demodulated, decoded into a data signal with decision-making, subjected to digital modulation, and followed with amplification and transmission. With a DF system, channel efficiency may be enhanced using the property of digital transmission as follows: the received signals transmitted from a base station and an MS are decoded separately into two data streams at a repeater. The two stream data (1,0) are modulo 2 added and the resultant data are transmitted from the repeater to the base station and the MS at the same time. On receiving and decoding the data, the base station and the MS add to the decoded data with data, which are sent from them and stored, to detect the data transmitted from their counterparts. The repeating system is called network coding, where the signal repeating in a downlink and an uplink is carried out simultaneously instead of independent repeating resulting in the enhancement of channel usage.

If functions of a repeater are enhanced maximally, it may be equivalent to a base station. Then it is called a layer 3 repeater, since it handles data of each user, for example, hybrid ARQ (Section 7.4.8). Repeaters of this kind are discussed in LTE systems (Section 9.7.6).

### 9.1.7 A Performance Analysis of Digital Cellular System

Performance of a cellular system is measured on some criteria as the maximum carried traffic accommodated per cell under a specified quality and a given system bandwidth. In a system focused on voice service, the maximum number of voice channels becomes a main issue. On the other hand, in a system designed for data services including voice over Internet protocol (VoIP), the maximum transmission rate (throughput), cell average transmission rates, and cell edge transmission rates are taken as a performance measure. Analyzing the performance of a cellular system is a tough task, since many matters spanning from physical layer to higher layers, such as radio wave propagation, voice coding, error control, modulation,

transmit/receive diversity, MIMO system, multiplexing, and channel scheduling. Therefore, the performance analyses are usually carried out by computer simulations (Section 5.6), which require a huge size programming and a long machine time even with an advanced computer. Here, we attempt to theoretically analyze the performance of a digital cellular system under much simplified assumptions.

**System Model** We assume a three-sectored hexagonal cell layout as shown in Figure 9.15. The system parameters are summarized in Table 9.2, partly according to [54].

The Rayleigh fading, the shadow fading, and the path loss are considered. MS or user equipment (UE) are uniformly distributed in a cell. Two-branch maximal combining diversity (Section 7.2.1) with no correlation between the branches is assumed. The correlation for shadowing loss from/to different base stations is considered.

Modulation methods are QPSK, 16QAM, and 64QAM, which are adaptively selected depending on the SINR value: if SINR is higher than 18 dB, 64QAM is used; if it is between 12 and 18 dB, 16QAM is used; and if it is 5–12 dB, QPSK is used. If the SINR takes a value <5 dB, the QPSK transmission speed is decreased proportionally to that true (not in dB) value. These SINR threshold values are decided

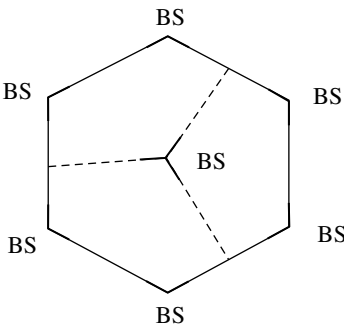


FIGURE 9.15 Cell layout.

TABLE 9.2 Assumed system parameters

Frequency band	2 GHz
Bandwidth	10 MHz
Propagation loss (dB)	$PL = 137 + 39\log_{10} d \quad (d : \text{km})$
Shadowing variance and correlation factor	$\sigma = 8 \text{ dB}, \rho = 0.5$
Transmit power: $P_T$	46 dBm, 40 dBm (BS), 23 dBm (UE)
Antenna gain: $G_B, G_U$	15 dBi (BS), 0 dBi (UE)
Diversity	2-branch maximal ratio combining
Noise figure	5 dB (BS), 9 dB (UE)
Noise power	−101 dBm (BS), −96 dBm (UE)
Cell configuration	Hexagonal 3 sector
Inter base station distance ( $d$ )	1.8 km

considering the error rate performance improvement with turbo decoding assuming that the same value of SNR and SINR shows the same error rate. Specifically, SINR=5 dB for QPSK is assigned for an error rate of  $10^{-3}$  with a turbo code ([60] in Chapter 7). Based on the SINR=5 dB for QPSK, the other required values for 16QAM and 64QAM are decided ([17] in Chapter 7).

Channel efficiency due to overhead is 80% and coding rate is 1/2. Then the transmission speed in 10MHz bandwidth becomes 8, 16, and 32Mbps for QAM, 16QAM, and 64QAM, respectively.

We assume for simplicity that three neighboring 120 degree sector cells interfere with each other: one BS sector cell at the center of Figure 9.1 and two BS sector cells that have areas overlapped with the center sector cell. The interference between different sector cells is approximated by interference between the two nearest neighboring cells.

**Channel Allocation** MSs that are located at cell edge suffer from signal quality deterioration due to received signal power attenuation and/or strong interference from other cells. To compensate for such deterioration of signal quality, transmit/receive diversity systems, interference suppression techniques with a sophisticated signal processing, and channel allocation scheduling are effective. Although these techniques can be applied for each pair of BS and its MS independently, coordinated operation of multiple BS (coordinated multipoint (CoMP)) becomes more effective (Section 9.1.5).

For a CoMP system, reduction of signaling traffic for coordinating multiple BS and of signal processing complexity becomes a key issue. From this viewpoint, a CoMP system is not favorable, which requires fast and heavy signaling between the BSs committed in coordination, as well as between a BS and an MS. Here, we present a method for CoMP, which uses a transmit and receive site-selection diversity system, and low-speed feedback signaling of channel quality information, that is, average (under fast fading) value of signal to interference and noise power ratio (SINR). When the cooperated transmission is always carried out, however, the total system throughput is decreased, since while a BS is using a channel, the other two BSs never use the channel. Here, a method for adaptive cooperation of base station is presented: SINR value at an MS is measured, and if the value is less than a given threshold, then the base station cooperation is carried out; else the base stations communicate with MSs independently.

When an MS is served at the cooperation mode using a channel, the channel is used exclusively for the MS, and there is no intercell interference. The downlink cooperation is done by transmitting signal in a channel to an MS from a selected BS, while the other two BSs never use the channel not to cause interference to that signal. The uplink cooperation is also done similarly.

**Analysis** Under the Rayleigh fading, received power is given as (Section 4.3)

$$p_R(q) = \frac{1}{b} e^{-q/b} \quad (9.2)$$

where  $b$  is average power. The distribution function of received power,  $\text{Prob}(q \leq q_0) \equiv P_R(q_0)$ , becomes

$$\begin{aligned} P_R(q_0) &= \int_0^{q_0} p_R(r) dr \\ &= 1 - e^{-q_0/b}. \end{aligned}$$

If we assume two-branch maximal ratio combining (MRC) diversity under the Rayleigh fading, the probability density function of received power becomes (Section 7.1.1)

$$p_R(q) = \frac{q}{b^2} e^{-q/b}. \quad (9.3)$$

Then the distribution function of received power is given using Equation 9.3 as

$$\begin{aligned} P_R(q_0) &= \int_0^{q_0} p_R(r) dr \\ &= 1 - e^{-q_0/b} \left( 1 + \frac{q_0}{b} \right). \end{aligned} \quad (9.4)$$

SNR distribution is given by replacing  $q_0$  by  $qN$ , where  $q$  is the SNR value and  $N$  is the noise power (which can be calculated from the noise figure and the noise bandwidth Section 2.2).

The cumulative distribution function  $P_R(q) = \text{Prob}(\text{SINR} \leq q)$  of signal to interference and noise power ratio (SINR:  $q = \gamma_1 / (\gamma_2 + N)$ ) under the Rayleigh condition is given as

$$P_R(q) = \int_0^{q(\gamma_2 + N)} d\gamma_1 \int_0^{\infty} p_1(\gamma_1) p_2(\gamma_2) d\gamma_2$$

where  $p_1(\gamma_1)$  and  $p_2(\gamma_2)$  are the probability density functions of received power for desired and interfering signals, respectively. Under the diversity reception, they are given by Equations 9.2 and 9.3, respectively. In this case, performing the integration we get,

$$P_R(q) = 1 - \left[ \frac{qN/b_2 + b_1/b_2}{q + b_1/b_2} + \frac{qb_1/b_2}{(q + b_1/b_2)^2} \right] e^{-qN/b_1} \quad (9.5)$$

We assume an MS is connected to a base station that shows the highest average power. This means that  $b_1 (\geq b_2)$  corresponds to the connected base station where the diversity reception is applied. In shadow fading, the logarithm of signal power,  $x (= 10 \log_{10} b)$ , is subjected to normal distribution as (Section 4.2)



$$p_s(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-x_m)^2}{2\sigma^2}\right) \quad (-\infty \leq x \leq \infty)$$

where  $x_m = \langle x \rangle$  and  $\sigma = \sqrt{\langle (x-x_m)^2 \rangle}$ . If two nonindependent signals in the shadowing are considered, the probability distribution function becomes (Section 2.2.7)

$$p_s(y_1, y_2) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left(-\frac{y_1^2 - 2\rho y_1 y_2 + y_2^2}{2\sigma^2(1-\rho^2)}\right)$$

where  $y_1 = x_1 - x_{1m}$ ,  $y_2 = x_2 - x_{2m}$  (suffix corresponds to each base station) and  $\rho$  is the cross-correlation coefficient. Since we assume a uniform distribution of MSs, probability density function of distance from a base station to an MS is given as

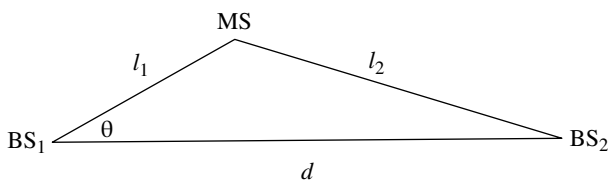
$$p_d(l) \cong \frac{2\pi l}{\pi R^2} = \frac{2l}{R^2} \quad (l = R_0 \sim R) \quad (R \gg R_0)$$

where  $R$  is the given cell radius and  $R_0$  is an arbitrary small distance introduced not to give an infinite power for  $l \rightarrow 0$ . The cell radius  $R$  is imaginarily chosen such that the total coverage area of the three interfering BSs becomes the same between the part of hexagonal area and the  $120^\circ$  three sectors, that is,  $\frac{\sqrt{3}}{2}d^2 = \pi R^2$ , where  $d$  is the distance between neighboring BSs.

The average power  $x_m$  (dBm) is given as follows:

$$x_m = \alpha - \beta \log_{10} l \quad (l : \text{in km})$$

where  $\alpha$  and  $\beta$  are given by the pass loss formula, transmit power, and antenna gains in Table 9.2, that is,  $\alpha = P_T + G_T + G_M - 136$  and  $\beta = 36$ . Distances to/from base stations depend on location of an MS as shown in Figure 9.16. When a distance between an MS and a BS is  $l_1$ , the distance  $l_2$  between the MS and the other BS depends on the angle  $\theta$ . For simplicity of numerical calculations, we neglect the effect of  $\theta$  on  $l_2$  and assume as  $l_1 = l$  and  $l_2 = d - l$ . (This assumption yields higher interference from BS. We confirmed by a numerical calculation that the assumption causes not significant



**FIGURE 9.16** Distances between base stations and a mobile station.

error in performance results.) Then average received powers in dBm at the MS (or at the BSs) are given as follows:

$$x_{1m} = \alpha - \beta \log_{10} l, \quad x_{2m} = \alpha - \beta \log_{10} (d - l)$$

where  $d$  is the distance between the two base stations.

When a base station is considered as an isolated sector cell, distribution function of SNR in the sector cell is given as

$$\begin{aligned} P_C(q) &= \int_{-\infty}^{\infty} p_S(x) dx \int_0^q p_R(q) dq \int_{R_1}^R p_d(l) dl \\ &= \int_{-\infty}^{\infty} p_S(x) dx \int_{R_1}^R P_R(q) p_d(l) dl \end{aligned}$$

The situation of intercell interference is different for downlink and uplink. In the downlink, when two base stations are operated independently corresponding to each MS, the SINR distribution in a sector cell is given by the following equation:

$$\begin{aligned} P_C(q) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_s(y_1, y_2) dy_1 dy_2 \int_{R_1}^R P_R(q) p_d(l) dl \\ &= \int_{R_1}^R P_R(q) p_d(l) dl \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_s(x_1 - x_{1m}, x_2 - x_{2m}) dx_1 dx_2 \end{aligned}$$

where  $P_R(q)$  is given by Equation 9.5, where  $b_1 = \max\{x_1, x_2\}$  and  $b_2 = \min\{x_1, x_2\}$ . For the cooperated mode of operation in the downlink,  $P_R(q)$  is given by Equation 9.4, where  $b = \max\{x_1, x_2\}$ .

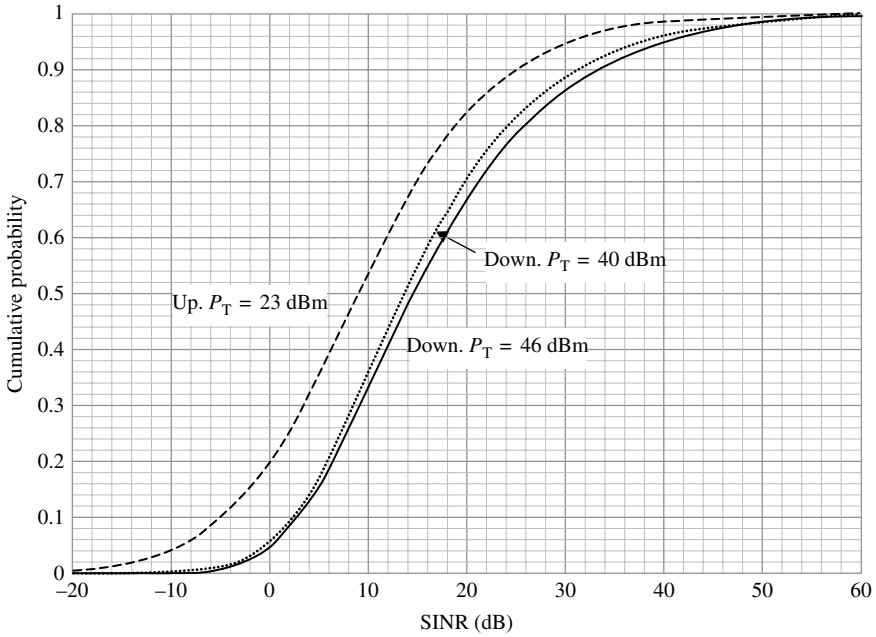
In the uplink, intercell interference situation becomes more complex since two MSs should be considered and the selection of a BS for a desired BS is not fixed, but depends on the propagation and the shadowing loss to establish a better connection. Here, we simplify this situation by assuming that the desired MS is always connected to a fixed cell but the best power received at the two candidate BSs is used. In this model, the SINR distribution is given as

$$P_C(q) = \int_{R_1}^R \int_{R_1}^R P_R(q) p_d(l) p_d(l_3) dl dl_3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_s(x_1 - x_{1m}, x_2 - x_{2m}) p_s(x_3 - x_{3m}) dx_1 dx_2 dx_3$$

where

$$x_{3m} = \alpha - \beta \log_{10} (d - l_3).$$

The expression of SNR distribution for the cooperative mode of operation in the uplink becomes the same as for the downlink.



**FIGURE 9.17** SINR distribution for independently operated base stations.

Figure 9.17 shows the distribution of SINR under the independent mode of operation. We can see the downlink performance is interference limited, since the SINR distribution is not improved even though the downlink transmit power is increased by 6 dB.

The cumulative distribution of SNR is shown in Figure 9.18, where two base stations are always cooperated. The downlink SNR distribution is greatly improved owing to avoidance of the intercell interference. For the uplink at a cumulative probability of 5%, SINR value of  $-9$  dB in Figure 9.17 is increased to  $-2$  dB (SNR) in Figure 9.18.

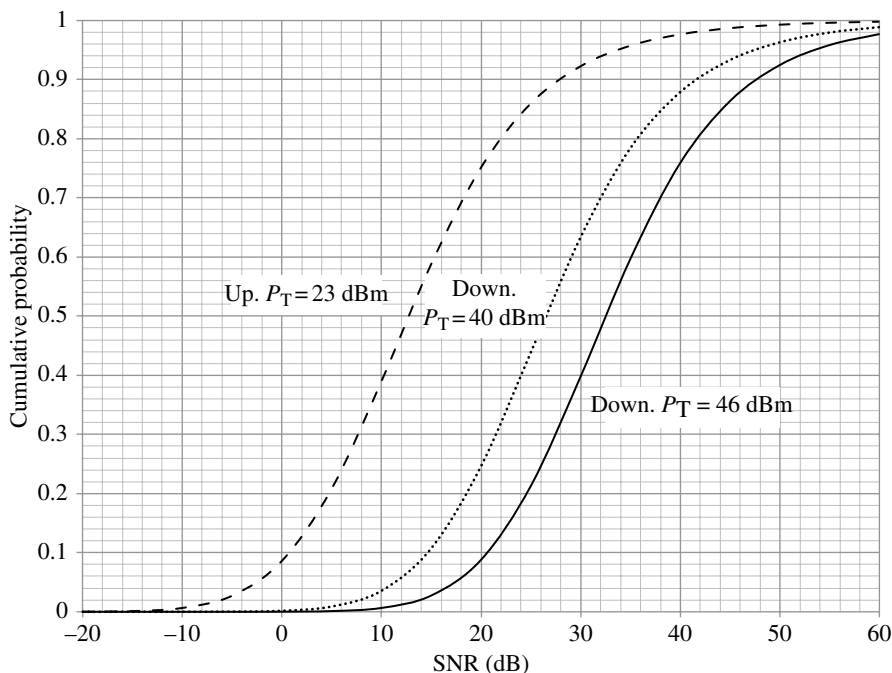
Figure 9.19 shows the SNR and SINR distribution in downlink with the adaptive mode of operation for transmit power of 40 dBm. Two base stations are cooperated adaptively when SINR value for an MS is dropped less than a given threshold value  $x_{th}$ . We can see that the probability of BS cooperation increases when the threshold value is increased and SNR and SINR distribution improves.

The uplink S(I)NR distribution for adaptive cooperation mode is shown in Figure 9.20.

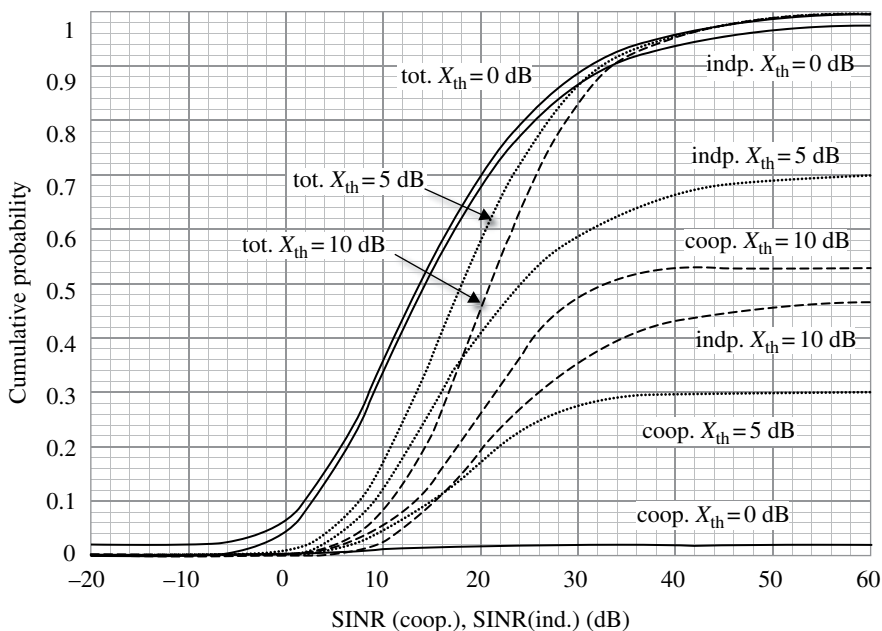
Throughput performance is calculated under the aforementioned assumptions. The average throughput  $G$  in Mbps is given from the SINR distribution as

$$G = 32Pr(64) + 16Pr(16) + 8Pr(4) + \Delta$$

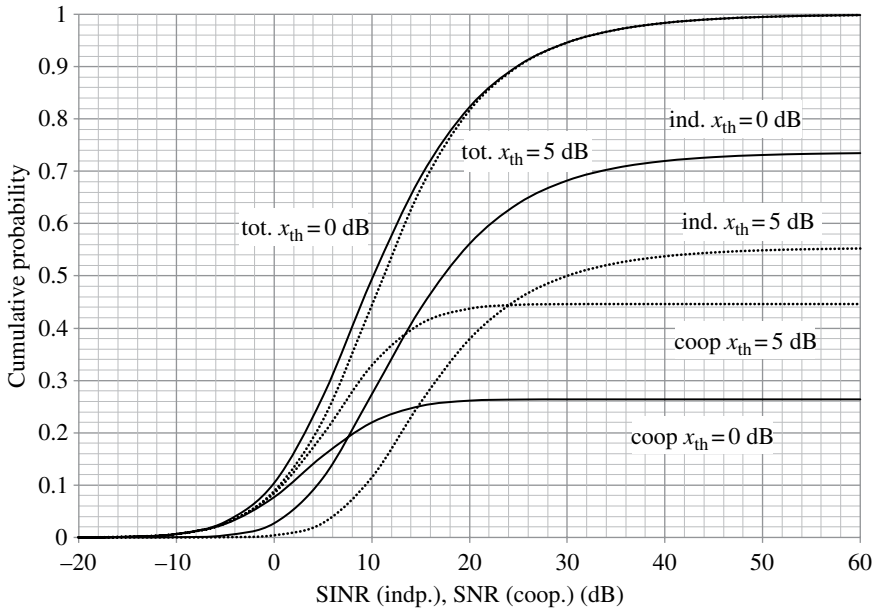
where  $Pr(64)$ ,  $Pr(16)$ , and  $Pr(4)$  denote the probabilities that 64QAM, 16QAM, and QPSK are selected, respectively.  $\Delta$  corresponds to the throughput by using QPSK with reduced rates as



**FIGURE 9.18** SNR distribution in a cell where two base stations are always cooperated.



**FIGURE 9.19** Downlink SNR and SINR distribution for adaptive cooperation mode of operation ( $P_T = 40$  dBm; Ind, independent; Coop, Cooperated; Tot = Ind + Coop).



**FIGURE 9.20** Adaptive mode uplink S(I)NR distribution.

$$\Delta = 8 \int_{-\infty}^5 10^{(x-5)/10} f(x) dx$$

where  $x$  is the S(I)NR value in dB and  $f(x)$  means the probability density function of  $x$ ;  $f(x)$  is given by differentiation of the cumulative distribution function of  $x$ .

Table 9.3 summarizes the cell average and cell edge 5% throughput performance for different modes of operation. The channel efficiencies in bps/Hz are given by dividing the throughput by 10.

For cooperative mode of operation, throughput is counted for the three cells as a whole: the average throughput per sector cell is given by dividing the throughput by 3. For independent mode of operation, the throughput is calculated independently for each cell. For adaptive mode of operation, let us denote calculated throughput by  $G_c$  with cooperative mode of operation, and that in independent mode by  $G_i$ , then the average throughput per sector cell is counted as  $G = G_c/3 + G_i$ .

When an MS locates closer to the cell edge, its S(I)NR value takes a lower value and therefore the attainable throughput decreases. As a measure of system performance, a throughput for an MS located near the cell edge is useful. We define a cell-edge throughput, where a cumulative probability of throughput lower than the cell edge throughput becomes 5%. For calculation of the 5% cell-edge throughput, we never take an average over the three cells: the throughput is counted as  $G = G_c + G_i$  to emphasize the maximum attainable throughput for a cell edge MS.

The all-time cooperation of base station shows a high cell edge 5% throughput, especially for downlink since the strong intercell interference in the downlink is

**TABLE 9.3 Summary of throughput performance for different modes of operation of base station**

	Average (Mbps)	5% cell-edge (Mbps)
Independent		
Down $P_T = 46\text{dBm}$	18	3.2
$P_T = 46\text{dBm}$	17	2.6
Up	13	0.32
Cooperated		
Down $P_T = 46\text{dBm}$	10	16
$P_T = 46\text{dBm}$	10	8.0
Up	6	1.6
Adaptive cooperated		
Down $P_T = 40\text{dBm}$ ( $x_{th} = 5\text{ dB}$ )	17	8.0
( $x_{th} = 10\text{ dB}$ )	16	8.0
UP		
( $x_{th} = 0\text{ dB}$ )	13	1.3
( $x_{th} = 5\text{ dB}$ )	12	1.4

avoided with the cooperated transmission. The increase of the cell edge 5% throughput in the uplink is owing to a site diversity effect in the noise-dominant uplink channel. However, it pays a penalty as the decrease of the cell average throughput.

The independent operation mode gives almost two times higher cell average throughput than the cooperated mode: the effect of the SNR distribution improvement with the base station cooperation is offset through the three cell averaging effect. The adaptive cooperation mode of operation enjoys a high cell edge throughput as the all-time cooperation mode, while the cell average throughput is only slightly decreased compared with the independent mode of operation.

With the adaptive base station cooperation method, MS that locates near to BS tends to be served at the independent mode, while MS far from BS is served with the cooperation mode. Thus, the reuse partitioning (Section 9.1.1) is automatically established.

The downlink cooperation is further enhanced by simultaneous transmission from two BSs. In this system, the received signal power at an MS is still subjected to the Rayleigh distribution with an average power, depending on coherent or noncoherent transmission as

$$b = \begin{cases} b_1 + b_2 & (\text{noncoherent}) \\ \left( \sqrt{b_1} + \sqrt{b_2} \right)^2 & (\text{coherent}) \end{cases}$$

where  $b_1$  and  $b_2$  are average received power from one of the two base stations, respectively.

The uplink cooperation can be expanded to simultaneous (instead of selection) reception at two BSs. When each BS uses the two-branch MRC diversity and output

signals from the BSs are combined with the MRC, the final received signal power density function is given by a convolution integral as

$$\begin{aligned} p_R(q) &= p_{R_1}(q) * p_{R_2}(q) \\ &= \int_0^q \frac{(q-x)}{b_1^2} e^{-(q-x)/b_1} \frac{x}{b_2^2} e^{-x/b_2} dx \\ &= \frac{1}{(b_1 - b_2)^3} e^{-(1/b_1 + 1/b_2)q} \left[ e^{q/b_1} ((b_1 - b_2)q + 2b_1b_2) + e^{q/b_2} ((b_1 - b_2)q - 2b_1b_2) \right] \end{aligned}$$

9.2 DIGITAL TRANSMISSION IN ANALOG MOBILE COMMUNICATION SYSTEMS

In analog mobile communication systems, digital signals must be transmitted for system control purposes. Table 9.4 [55] shows a summary of digital transmission technologies used in analog mobile communication systems such as cellular systems for NTT, US (AMPS), the Nordic (NMT), and the two-way radio system (MCA) in Japan. The data transmission speed is slow. Major parts of circuits including an FM modulator and a limiter–discriminator are used in common between the digital transmission and the analog FM voice signal transmission. Digital data is transmitted

TABLE 9.4 Digital transmission technologies in some analog mobile communication systems [55]

	Mobile telephone system			Trunked system
	NTT	AMPS	NMT-450	MCA <sup>a</sup>
Bit rate (bps)	300	10,000	1200	1200
code	Split phase	Split phase	NRZ	NRZ
Modulation	FSK	FSK	Subcarrier FSK <sup>b</sup>	Subcarrier FSK <sup>b</sup>
Maximum frequency deviation (kHz)	±4.5	±8	±3.5	±3.5
Error correction	BCH (recycled transmission)	BCH (recycled transmission)	Hagelberger	Hagelberger
Channel spacing (kHz)	25	60	25	25
Frequency band (MHz)	800	800	450	800

Reproduced by permission of Copyright © 1985, IEICE.

<sup>a</sup>Press-to-talk system for private communication.

<sup>b</sup>Mark frequency = 1200 Hz; space frequency = 1800 Hz.

through analog voice channels, where the dc component is blocked. Therefore, Manchester coding or subcarrier systems are used (Section 3.2).

9.3 PAGING SYSTEMS

A paging system is a one-way (wire-line to mobile) digital transmission system. Initially, tone-only systems appeared, where an audio tone is generated when a terminal is called. The next-generation systems transmit numerical signals to be displayed at a receiver. Even messages can be sent in modern paging systems. Message paging systems had grown with evolution of the electronic-mail services. The message communication hardly interrupts the person being called, and a vibrator mode of a call alerting gives no interference to other people.

The pager terminals have become compact, lightweight, and slim. Direct conversion systems are employed for some card-type receivers to reduce the size of the terminal.

The Post Office Code Standardization Advisory Group (POCSAG) systems are widely known [56]. The specification is summarized in Table 9.5 and signal format is shown in Figure 9.21. For a tone-only service, 230,000 subscribers can be accommodated per channel.

To cope with increasing numbers of users and roaming within Europe, the European Telecommunications Standards Institute (ETSI) standardized a paging system called European Radio Messaging System (ERMES) [57]. The specification is shown in Table 9.6. The channel data rate is increased to 6250 bps and multilevel FM is employed

TABLE 9.5 POCSAG paging system specifications

Frequency	280 MHz band
Channel spacing	25 kHz
Modulation	NRZ-FSK
Channel speed	512 bps
Error correction code	BCH (31,21)

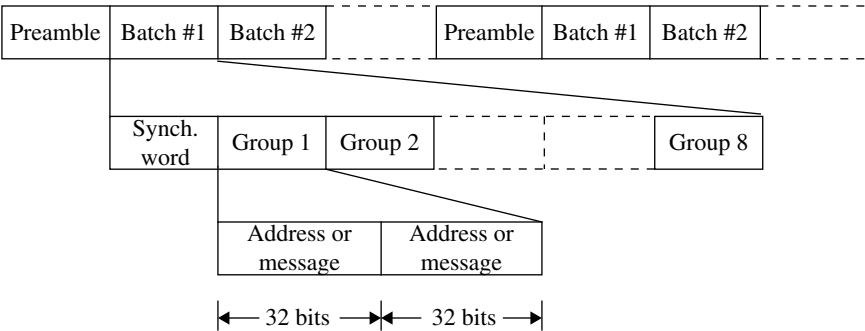


FIGURE 9.21 Signal format for POCSAG paging system.



**TABLE 9.6   ERMES specifications**

Code word type	Shortened cyclic code (30,18)
Address capacity	32 million for each country
Frequency band	169.425–169.8 MHz
Number of RF channel	16
Channel spacing	25 kHz
Modulation	4PAM/FM (4-level FSK)
System bit rate	6250 bps

to obtain high-speed transmission within the channel spacing of 25 kHz. Simultaneous transmission from multiple base stations is used. In this system, signals from different base stations should be time synchronized within 50 ms. ERMES provides tone, numeric, alphanumeric, and transparent data paging services. Supplementary services, such as roaming, message repetition, legitimization, urgent message indication, and so on, are included. Pager terminals must be able to scan through 16 RF channels.

In Japan, an advanced radio paging system was standardized based on a system called FLEX\* (\*registered trademark of Motorola). Three data rates of 1600, 3200, and 6400 bps can be selected to adapt to the service requirements: a data rate is chosen by compromising the traffic density and the coverage area size; a high data speed service is used for a high subscriber density area with a shortened maximum service range. Four-level FSK is used for data rates of 6400 and 3200 bps and 2-level FSK for data rates of 3200 and 1600 bps. A time diversity system where the same data are sent up to four times is adopted to obtain long service ranges. A code word consists of BCH (31,21) and a parity bit. The flexible operation of the system is performed through sending a frame information signal that tells data rates, number of repeated transmissions, and so on. A variety of services such as tone-only, numeric and alphabetic message, and data transmission are efficiently accommodated through the use of different fields (i.e., address, vector, and message fields). The vector field data indicates the start and end positions of the message for a called terminal.

**9.4   TWO-WAY DIGITAL MOBILE RADIO**

The so-called two-way mobile radio is a press-to-talk private radio system for applications such as car dispatch services and police communications. The first application of digital voice transmission in mobile radio communications was the two-way radio system. Since digital mobile radio technologies were not well developed at that time, adaptive delta modulation (ADM) voice coding and digital FM were used [58]. A 150 MHz digital mobile radio equipment using 16 kbps ADM voice coding and four-level FM at 25 kHz channel spacing was reported [59, 60]. The RF, IF, and modulator/demodulator circuits are used for both analog and digital modes of communication.

A digital mobile radio for police service was reported in Ref. [61]. Using efficient voice coding of 8 kbps and constant envelope digital modulation, this system is operated at 12.5 kHz channel spacing with an error rate of  $5 \times 10^{-2}$ .

9.5 MOBILE DATA SERVICE SYSTEMS

Mobile radio communication services such as dispatch traffic were improved with the introduction of digital data transmission systems: transmission of text and other data messages can replace or assist voice communication. These kinds of digital mobile communication systems described in the following had been in commercial use. However, services are terminated with evolution of digital cellular systems.

9.5.1 MOBITEX

MOBITEX was a mobile communication system [62, 63] developed by the Swedish Telecommunications Administration, where digital data as well as analog voice signals are transmitted through trunked frequency channels. The base stations operate in duplex, while the mobile units operate in two-frequency simplex or duplex. The channel spacing is 25 kHz. The digital modulation is FFSK with a data rate of 1200 kbps. A BCH (15,10) code for data coding and ARQ are used in the radio channel.

9.5.2 Teleterminal System

Teleterminal system was a public land mobile radio data communication system developed in Japan [56, 64]. Data signals are transmitted in trunked packet-switched channels. The summary of the system specifications is shown in Table 9.7. The Reed–Solomon code (15,11,5) and ARQ are adopted for error control in the radio channel. The multiple-access method is a polling system, where a polled terminal can transmit data packets.

9.5.3 Mobile Data Systems in Analog Cellular Systems

Since cellular systems are connected to the wire-line public switched telephone network (PSTN), it was desired to extend data services in the PSTN to existing analog cellular mobile systems. To this end, the cellular data link control (CDLC) system [65] and the cellular digital packet data (CDPD) system [66] were proposed.

TABLE 9.7 Teleterminal system specifications

	Mobile	Base
Frequency	800 MHz band	
Mode of radio transmission	Two-frequency simplex	Two-frequency duplex
Modulation	FSK	
Channel spacing	25 kHz	
Channel speed	9600 bps	
Transmit power	5 W	20 W
Transmit/receive spacing	55 MHz	

In the CDLC system, the CCITT V26 modem is used for a baseband voice channel spanning from a PSTN to an analog FM cellular system. BCH (16,8) and Reed–Solomon (72,68) codes are selected for the forward channel depending on the circuit conditions, and the Golay (23,12) code is used in the backward channel. The protocol is a full duplex with layer 2 protocol based on HDLC standards.

In the CDPD system, data signals are transmitted as packets through a full-duplex radio channel with channel data rates of 19.2 kbps. The mobile data base stations are probably cosited with analog FM cellular base stations. The CDPD system detects the cellular system channels that are not used for voice communication and uses them for data transmission. When a voice communication starts in the channel, it stops the data communication in that channel and switches it to another idle channel (channel hopping). The channel hopping is controlled by the mobile data base station broadcasting the channel hopping signal to the data terminals.

The CDPD protocol supports the physical layer and the data link layer. Any of the upper-layer protocols can be used: the transmission control protocol/Internet protocol (TCP/IP) and the Open System Interconnection (OSI) [28] for computer networks. Through the use of the value-added networks, various data services were provided for mobile data users.

## 9.6 DIGITAL CORDLESS TELEPHONE

Analog cordless telephones are widely used all over the world. Although mobility is limited to a distance of several tens of meters, many people enjoy the wireless communication. The penetration is limited to in-house use. The digital cordless telephone system is expected to be widely used in business and public areas as well as at home, owing to high capacity, high security due to digital scrambling, and improved data service capability. Considering many different applications of digital cordless telephone, the performance is limited by the air interface. In this section, proposed digital cordless telephone systems are described.

### 9.6.1 Second-Generation Cordless Telephone

The second-generation cordless telephone (CT-2) was developed in the United Kingdom. The specification is summarized in Table 9.8 together with those for other systems. Time-division transmit and receive duplexing (TDD) on a frequency was adopted. Since one block of frequency band is sufficient for this system, in contrast to the two-frequency duplexing system, the spectrum allocation becomes easier. In addition, diversity transmission is realized by employing circuits for the diversity system only at the base station due to high correlation between the uplink and downlink channels on the same frequency. Furthermore, direct communication between mobile terminals is easily made.

The FDMA system is adopted with transmission rates of 32 kbps. The voice coding is 32 kbps ADPCM. No channel coding is used. These concepts are based on emphasizing low system cost and high voice quality.

**TABLE 9.8 Digital cordless telephone specifications**

	CT-2	DECT	PHS
Frequency band (MHz)	900	1900	1900
Access system	FDMA/TDD	TDMA/TDD	TDMA/TDD
Channels/carrier	1	12	4
Voice coding	32 kbps ADPCM	32 kbps ADPCM	32 kbps ADPCM
Modulation	GMSK	GMSK	$\pi/4$ QPSK
Channel speed (kbps)	72	1152	384
Carrier spacing (kHz)	100	1728	300
Frame length (ms)	2	10	5
Transmit power (mW)			
(peak)	20	250	80
(average)	10	10	10

A public service known as the Telepoint with CT-2 was launched first in the United Kingdom. This system is limited to calls originating from the MS: a paging system is used for calling an MS. This fact and the smaller service area than with cellular phone systems are the probable reasons for lack of success of the Telepoint service in the United Kingdom.

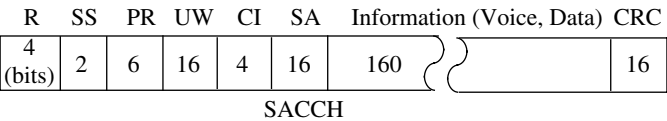
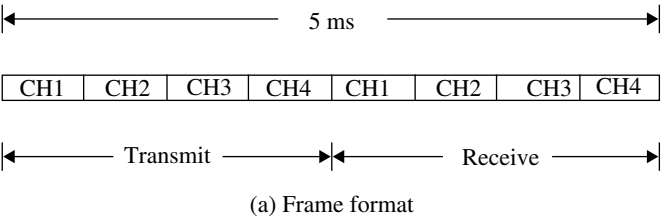
### 9.6.2 Digital European Cordless Telecommunications

Digital European Cordless Telecommunications (DECT) was standardized by the ETSI. It is a 12-channel per carrier TDMA/TDD system with channel speed of 1152 kbps. It uses frequencies in the 1900 MHz band. The other major specifications are the same as those for the CT-2 system. Rather high channel data rates will characterize the services provided by the DECT system.

### 9.6.3 Personal Handy System

The Personal Handy System (PHS) was standardized in Japan. Initially, it was called Personal Handy Phone (PHP). It is a 4-channel per carrier TDMA/TDD system. The modulation is  $\pi/4$  shifted QPSK. The frame length is 5 ms. The other major specifications are similar to those for DECT. The frame structure is shown in Figure 9.22.

The peak transmit power is 80 mW, resulting in an expected zone radius of 100–200 m. Functions for interference detection before channel selection as well as during communication are required as part of the standardized functions. For starting communication, a radio channel that is proven not to interfere with other zones by performing the interference detection is selected at each base station. During communication, if interference is detected, the channel is switched to another channel (intracell hand-off).



- R : Ramp time
- SS : Start symbol
- PR : Preamble for bit synchronization
- UW : Unique word for frame synchronization
- CI : Channel identification
- CRC: Cyclic redundancy check
- SA : Slow associated control channel

(b) Traffic channel format

FIGURE 9.22 PHS signals.

9.7 DIGITAL MOBILE TELEPHONE SYSTEMS

The fast growth of digital mobile radio communication technologies is strongly driven by research and development activities for digital mobile telephone systems. A pioneering experimental TDMA system was reported in 1982 [67], recognizing the advantages of a TDMA approach. However, the field test results showed difficulty in high-speed TDMA digital transmission due to the fast frequency-selective fading effects. The first Nordic Seminar on Digital Land Mobile Radio Communications in 1985 astonished digital mobile radio communications researchers, or at least the author of this book, by presenting so many ambitious proposals, some of which affected the later pan-European (Group Special Mobile (GSM)) digital cellular systems. The pan-European organization for the GSM system standardization was launched. It was not long before efforts for standardization of digital mobile telephone systems started in the United States and Japan. In this section, the GSM, the American digital cellular, and the (Japanese) personal digital cellular systems are described in brief. For details, refer to Refs. [2–13].

### 9.7.1 The GSM System

The European countries set up a committee called GSM for making a standard pan-European digital mobile telephone system that could be introduced in the 1990s. Criteria for the GSM (later called Global System for Mobile Communications) systems are as follows:

- Spectral efficiency
- Subjective voice quality
- Mobile cost
- Hand-portable feasibility
- Base station costs
- Coexistence with the existing systems.

A frequency band (890–915 MHz for mobile, 935–960 MHz for base) that had not been used for analog systems in many European countries was assigned.

Many systems were proposed as candidates for the GSM system. It is worth recalling them here. Table 9.9 summarizes the proposed systems. All the multiple-access methods FDMA, TDMA, and CDMA are included. An automatic equalizer is assumed to be used in the TDMA systems. Through the efforts of many people, the GSM system was standardized. Table 9.10 [56] summarizes the GSM system specifications.

A TDMA system was selected by a comparison between analog versus digital, FDMA versus TDMA, and narrowband versus wideband TDMA. The regular pulse excited predictive coding with long-term prediction (RPE-LTP) with a voice coding rate of 13 kbps was adopted (Section 7.7.5). The block length for voice coding is 20 ms. The coded voice signal is protected with a convolutional code resulting in a gross rate of 22.8 kbps. For digital modulation, DPM was tentatively chosen; and later, GMSK with baseband filtering, where the 3 dB bandwidth normalized by the bit rate  $B_b T$  is 0.3, was selected. This digital modulation scheme is a constant envelope

**TABLE 9.9 Proposed systems for the GSM system**

System	Company	Multiple access	Modulation	Bit rate (kbps)	Channel spacing (kHz)
S900D	Bosch/Ant/Matra	TDMA	4-level FSK	128	250
DMS90	Ericsson	TDMA	GMSK	340	300
SFH900	LCT/TRT	TDMA/CDMA	GMSK	200	150
CD900	SEL/AEG/ATR/SAT/ITALTEL	TDMA/CDMA	QPSK	4000	6000
MATS-D	TeKaDe/TRT		QAM GTfM	1218 19.5	1250 25

modulation that is different from the linear modulation that is used in the US and the Japanese mobile telephone systems. It is not clear whether linear modulation was discussed as a candidate method. The proposal for linear modulation for mobile communication appeared in 1987 [68].

The TDMA frame structure is shown in Figure 9.23. A TDMA frame includes eight time slots. The time slots for the uplink are intentionally delayed relative to those for the downlink (base to mobile) for use of a switch instead of a duplexing filter for transmit/receive duplexing purposes.

TABLE 9.10 GSM system summary

Frequency band	890–915 MHz (mobile to base) 935–960 MHz (base to mobile)
Channel spacing	200 kHz (interleaved)
Number of frequency channel	124
Multiple access	8 channel-TDMA Time-slot length 577 $\mu$ s (156.25 bit)
Modulation	GMSK (BbT = 0.3)
Channel speed	270.833 kbps
Voice coding	RPE-LTP (13 kbps, 22.8 kbps with FEC)
User data speed	9.6, 4.8, 2.4 kbps
Automatic equalizer	Capable of the maximum delay spread of 16 $\mu$ s
Frequency hopping	217 hops/s
Base station coverage (radius)	0.5–35 km
Transmit power	0.8–20 W (mobile), 2.5–320 W (base)

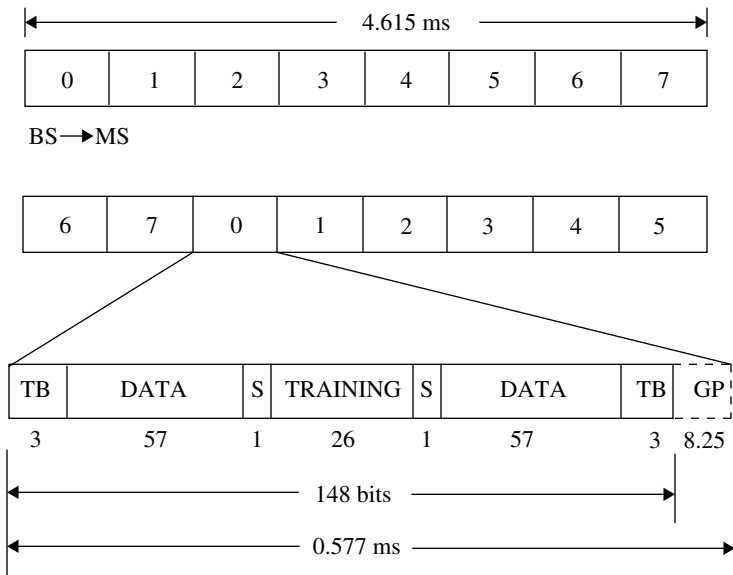


FIGURE 9.23 GSM system frame format.

The transmission rates become 270.833 kbps, accommodating eight channel-coded time slots. This transmission rate cannot be achieved for a mobile telephone system unless an automatic equalizer is used. As a standardized performance of the equalizer, the (average) delay spread (Section 4.4) of  $16\mu\text{s}$  must be coped with. The standardization for the equalizer is not necessary; the choice of equalizer is left for the equipment manufactures. Viterbi or decision feedback equalizers (Section 7.3) are used for this purpose. The training signals of length of 26 bits are used for equalizer training as shown in Table 9.11. The autocorrelation function of the training signal is shown in Figure 9.24. This is calculated by taking correlation between the training signal code and the central 16 bits of the code, assuming  $-1$  for  $0$ . The autocorrelation function shows a sharp peak at the center and a null with  $\pm 5$  symbol time. Thus, with this code, a channel impulse response with a time duration within  $\pm 5$  symbols results.

The GSM transceiver must be frequency agile for two purposes: (i) for an MS to monitor one channel on different frequencies during a frame length of 4.6 ms including the transmit and reception periods and (ii) to randomize cochannel interference. The base station hops at every frame slot, resulting in a hopping rate of 217 hops/s.

TABLE 9.11 Training signal codes of the GSM system

Training sequence no.	Values
1	00100101110000100010010111
2	00101101110111100010110111
3	01000011101110100100001110
4	01000111101101000100011110
5	00011010111001000001101011
6	01001110101100000100111010
7	10100111110110001010011111
8	11101111000100101110111100

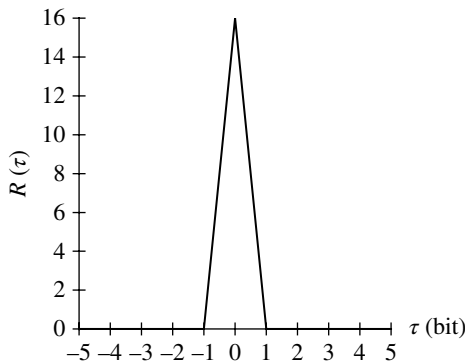


FIGURE 9.24 Autocorrelation function of a training signal for GSM systems.



Channels are logically divided into traffic channels (TCHs) and control channels. The TCH is an information-bearing channel. The control channels include the following:

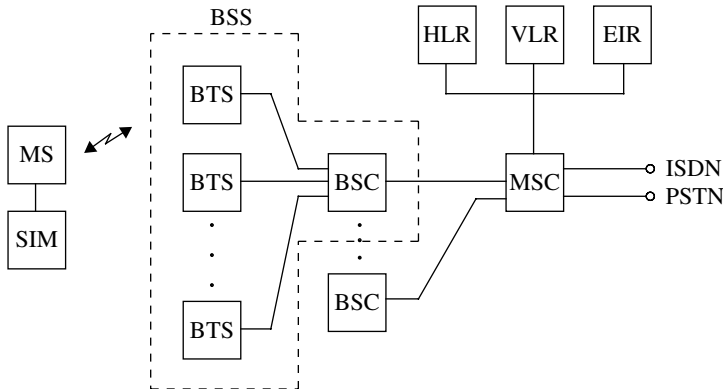
- Broadcast control channels (BCCHs) are for broadcasting information about the system to all MSs.
- Frequency correction channels (FCCHs) are for MSs to adjust their frequency to a stable frequency of a base station.
- Paging channel (PCH) is for paging MSs.
- Random access channels (RACHs) are uplink channels, where MSs make random access to the channels, in order to send control signals (e.g., channel demand for call origination at MSs).
- Slow associated control channels (SACCHs) are channels for transmitting slow control signals together with information signals.
- Fast associated control channels (FACCHs) steal traffic time slots to be used for fast control requirements such as hand-off.

The GSM network configuration is shown in Figure 9.25. Some base station controllers are connected to paging, hand-off, and so on. An MS can be used between different people by use of the subscriber identity card. Mobile switching centers are equipped with file systems such as home location register, visiting location register, and equipment identity register. Many data services as well as voice service are provided with the standardized protocols, which are beyond of the scope of this book.

The second phase specifications are standardized in June 1993. New services are added and a frequency band is attached below the frequency band of the first phase. Together with the second phase specifications, a 1800 MHz system called DCS1800 is standardized as personal communication networks. A half-rate voice coding system is also standardized in 1995, where TDMA frames are used in every other frame for a coded voice data stream. The GSM became a technical committee named Special Mobile Group (SMG) in the ETSI in 1988. The GSM has widespread all over the world as its later name expresses.

### 9.7.2 Digital Cellular Systems in North America

In the United States, the shortage of capacity of the existing analog system (AMPS) was expected with the increase in subscribers, especially in big cities such as New York, Chicago, and Los Angeles. To cope with this, efforts toward a higher-capacity system were begun. A condition was assumed that no new frequency spectrum was assigned to this system, but that the (analog) AMPS system should be gradually replaced with the new system. The Telecommunications Industry Association (TIA), which was the main body for the new system standardization, decided to adopt a TDMA system in 1990. In 1991, the TIA issued the standard system (IS-54). A dual-mode (analog and digital) system was considered owing to the requirement for gradual replacement.



- MSC : Mobile service switching center  
 HLR : Home location register  
 VLR : Visitor location register  
 EIR : Equipment identity register  
 BSS : Base station system (BTS + BSC)  
 BSC : Base station controller  
 BTS : Base transceiver station  
 MS : Mobile station  
 SIM : Subscriber identity module  
 ISDN : Integrated services digital network  
 PSTN : Public switched telephone network

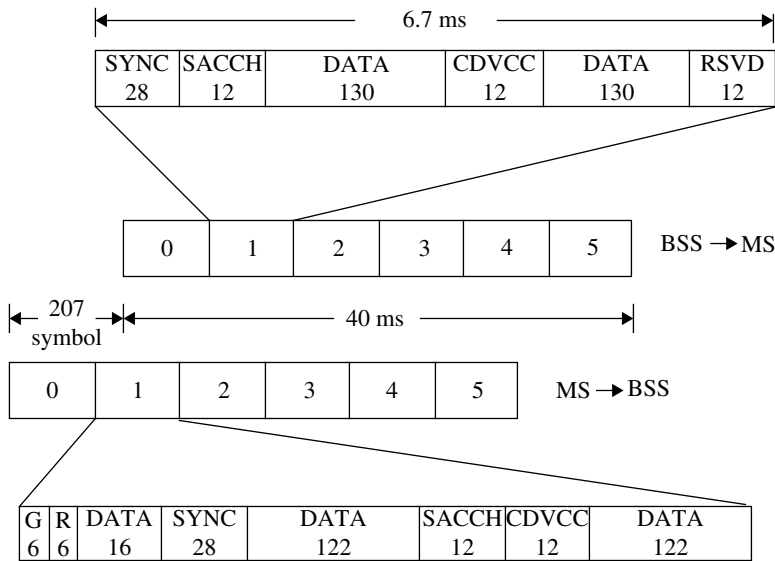
**FIGURE 9.25** GSM network configuration.

Table 9.12 summarizes the system specifications [56]. The interleaved channel spacing of 30kHz is the same as that for the AMPS system for ease of the dual mode of operation. A 3-channel TDMA with six-time slot system is adopted. The frame structure is shown in Figure 9.26. A voice signal uses slots 0 and 3 or 1 and 4 or 2 and 5. If half-rate voice coding is adopted, it becomes a 6-channel TDMA system. For voice coding, VSELP is used with 7.95 and 13.5kbps without and with error protection, respectively. The digital modulation is  $\pi/4$  shifted QPSK, which was the linear modulation first introduced into cellular systems. A maximum delay difference of 40 $\mu$ s in radio channels must be handled with an automatic equalizer. A long slot length of 6.7ms requires a high tracking speed for the equalizer, since the channel characteristics change significantly during the burst time length under fast fading.

The IS-54 system has three times the capacity of the AMPS system, if the reuse factor is the same between the two systems. Discussion of the reuse factor is difficult

TABLE 9.12 American digital cellular summary

Frequency band	869–894 MHz (BSS to MS) 824–849 MHz (MS to BSS)
Channel spacing	30 kHz (interleaved)
Number of frequency channel	832
Multiple access	3-channel/6-time slot TDMA
Time-slot length	40/6 ms
Modulation	$\pi/4$ -shifted QPSK
Channel speed	48.6 kbps
Voice coding	VSELP (7.95 kbps, 13 kbps with FEC)
Automatic equalizer	Capable for maximum delay difference of 40 $\mu$ s
Base station coverage (radius)	0.5–40 km
Transmit power	0.6–4 W (mobile)



- SYNC : Synchronization and training
- DATA : User information or FACCH (fast associated control channel)
- CDVCC : Coded digital verification color code
- RSVD : Reserved
- G : Guard time
- R : Ramp time
- SACCH : Slot associated control channel

FIGURE 9.26 Frame structure for the American digital cellular (IS-54) system.

since it is dependent on a required protection ratio for cochannel interference, the probability that the signal-to-noise ratio drops below a given threshold, and the voice quality under cochannel interference and fading conditions. For comparison of voice quality between different voice coding systems including an analog FM system, the test of opinion of many users must be carried out in an actual communication channel environment.

After the standardization of the IS-54 TDMA system, another candidate system using a CDMA architecture [69] was proposed by Qualcomm Inc. This CDMA system was also adopted as another standard (IS-95) system. The specifications are summarized in Table 9.13. It is a direct sequence spread spectrum (DSSS) system. The multiple-access system is DS CDMA, which is the first SS system standardized for cellular systems. The spreading factor is 128. The spreading code is a combination of pseudo noise (PN) codes and Walsh functions. By using the Walsh functions with length of 64 bits, 64 orthogonal channels are provided for a base station. By using different PN codes between base stations, the same frequency is used for each base station, that is, the reuse factor or cell group is 1. A spectrum efficiency of around 10 times that of the AMPS system is claimed. For this result, in addition to the one-frequency reuse, automatic transmit power control, soft hand-off or macrodiversity, voice-activated transmission, and cell sectorization are assumed. Automatic transmit power control, to maintain received signal level within a given range at a base station, is necessary for a spectrum-efficient CDMA system to cope with the near-far effect (Section 4.5). Soft hand-off means that the hand-off is gradually made as in macrodiversity between base stations. This technique is effective under shadowing conditions. Some channels must be reserved for soft hand-off purposes.

Voice-activated transmission, where the carrier signal is transmitted only when voice signal is detected, decreases the interchannel interference by the ratio of silence time period to the whole speech period.

In this system, sectorized cells are also considered. The effect of sectorization results in an  $N$  times increase in spectrum efficiency by reducing the size of the service area to  $1/N$ , where  $N$  is the number of sectors.

Frequency guard bands must be reserved between different frequency carriers, which reduces the spectrum efficiency.

**TABLE 9.13 CDMA digital cellular system (IS-95) summary**

Spread spectrum system	Direct sequence
Chip rate	1.2288 MHz
Data rate	9600 bps
Spreading factor	128
Modulation	QPSK
Bandwidth	1.23 MHz
Voice coding	QCELP (8 kbps)
Error correction	Convolutional coding/Viterbi decoding
Automatic transmit power control	Dynamic range of 85 dB
at mobile station	Step size of 0.5–1 dB

Thus, the high capacity of the CDMA system is maintained through many advanced techniques. Validation of the claimed capacity of the CDMA system is difficult because all the MSs must in principle take part in the experiment. In contrast, interchannel interference experiment for TDMA or FDMA systems can be carried out using a small number of MSs that interfere with a considered channel.

9.7.3 Digital Cellular Systems in Japan

The Japanese digital cellular standard was made by the Research and Development Center for Radio Systems (RCR) in April 1991. The system was called Personal Digital Cellular (PDC). A summary of the specifications is shown in Table 9.14 [70]. New frequency bands are assigned at 800/900 MHz and 1.5 GHz. The carrier spacing of 25 kHz is a little narrower than that for the US IS-54 system of 30 kHz. This is achieved with voice coding at a lower rate of 11.2 kbps. The modulation and multiple-access system are the same as those for the US IS-54 system. The half-rate voice coding was already standardized: the PSI-CELP (Section 7.7.6) at a coding rate of 3.45 and 5.6 kbps without and with error protection was adopted.

The automatic equalizer and diversity system are optional. Since a diversity receiver shows immunity against frequency-selective fading as well as flat fading, the automatic equalizer might be unnecessary as long as the delay spread is not too high. Under Rayleigh fading, the diversity system requires an average carrier signal ratio (CIR) of 11 dB for a bit error rate of  $2 \times 10^{-2}$ , which is the threshold value for the system; without the diversity system, a CIR of 16 dB is required.

The PDC had a great success as a cellular business owing to introduction of new data services and low-cost and compact base stations with a common linear power amplifier with adaptive nonlinearity compensations (Section 8.6.3). After the appearance of the W-CDMA system, PDC lost their subscribers and stopped new application in November 2008 and ended its service in March 2012.

TABLE 9.14 Specification summary for the Japanese digital cellular system

Frequency band	810–826 MHz (BSS → MS) 940–956 MHz (MS → BSS) 1477–1489 MHz (BSS→ MS) 1429–1441 MHz (MS → BSS) 1501–1513 MHz (BSS →MS) 1453–1465 MHz (MS → BSS)
Carrier spacing	25 kHz (interleaved)
Modulation	$\pi/4$ shifted QPSK
Multiple access	3-channel TDMA (full rate) 6-channel TDMA (half rate)
Channel speed	42 kbps
Voice coding	VCELP (6.7 kbps, 11.2 kbps with error protection) PSI-CELP (3.45 kbps, 5.6 kbps with error protection)
Automatic equalizer	Optional
Diversity	Optional

### 9.7.4 Evolution of the Second-Generation Systems

The digital cellular systems described so far are sometimes called as the second-generation system: the first-generation system corresponds to an analog system. The second-generation system had succeeded in the world. A reason for the success is its capability of data communication as well as voice services. E-mail and Internet access services through a cellular phone became possible. For an example, the so-called i-mode, which is an Internet access service provided by NTT Docomo in a PDC system, gained a great success owing to its well-designed business model attracting many users even who had no experience of wire-line Internet services.

The GSM evolved into the phase 2+ system. This system includes the circuit-switched data service (HSCSD) with the maximum speed of 57.6kbps and the general packet radio service (GPRS) with the maximum speed of 115.2kbps. Four time slots of a GSM signal are combined and used for a user to enhance the transmission speed. To evolve GSM to a third-generation system, enhanced data rates for global evolution (EDGE) was standardized, where 8PSK took the place of GMSK as its digital modulation scheme. The maximum transmission rate is 384kbps. To enhance the transmission rate of EDGE, EDGE evolution was added in release 7 by a standardization group third-generation partnership project (3GPP). The maximum use of time slots is expanded up to 10 using two frequencies. QPSK, 16QAM, and 32QAM are introduced as digital modulation scheme. Furthermore, the turbo codes and receive diversity at a mobile terminal are introduced, and transmission delay time is decreased to 100ms. The maximum transmission rates are 1.3Mbps and 653 kbps in the downlink and the uplink, respectively.

The IS-54 system evolved as IS-136. The major modifications are extension of a battery life, use of 1800MHz as well as 800MHz bands, and introduction of digital control channels. As a voice coding system, IS-641 enhanced full rate (EFR) with ACELP (Section 7.7.6) was added to the conventional VCELP. This system has a voice quality comparable as PCM in a wire-line channel. The IS-136 system was taken the place of GSM and its services were terminated in 2007 or 2008.

The IS-95 system is called as CDMA One or cdmaOne. IS-95 evolved first to IS-95A with an enhanced transmission rate of 14.4kbps. Then the second evolution system, IS-95B, attained a transmission rate of 115.2kbps enhanced with aggregation of eight data channels.

### 9.7.5 The Third-Generation System

Following the success of the second-generation systems and to proceed further evolutions, investigations for next-generation (3G) systems were started under a name of Future Public Land Mobile Telecommunications System (FPLMTS) by International Telecommunication Union (ITU) since around the mid-1980s. Later, the name changed as International Mobile Telecommunications-2000 (IMT-2000). Its concepts are aimed to build wireless infrastructures where various services are provided with qualities as comparable to those with wire-line communications in

different wireless transmission environments, and different wireless systems are integrated seamlessly. Specifically, services such as mobile TV phone and high-speed Internet access become realizable by increasing transmission rates, and a user can receive the services even if he/she moved to any place in the world (international roaming). The allocated frequency bands are 1885–2025 and 2110–2200 MHz. For satellite communication services, 1980–2010 and 2170–2200 MHz bands were allocated. The minimum transmission rates for a user are defined as

- Vehicular environments: 144 kbps
- Pedestrian environments: 384 kbps
- Indoor environments: 2.048 Mbps

Ten systems were proposed from the United States, Europe, Japan, Korea, and China. Table 9.15 shows standardized 3G systems. These systems are categorized into three groups: wideband version of CDMA systems, evolved TDMA systems, and combined CDMA and TDMA systems. The CDMA system includes IMT-DS (direct spreading) with a SC and IMT-multicarrier (MC) with three carriers. The former pursues the benefits of new wideband systems. The latter aims to succession of the existing CDMA systems. The IMT-DS and IMT-MC are proposed from the groups called 3GPP and third-generation partnership project 2 (3GPP2), respectively.

The IMT-DS is a direct spreading system with 5 MHz bandwidth, which was proposed as W-CDMA from Japan and as UTRA-FDD from Europe. Its major specifications are shown in Table 9.16. By increasing the bandwidth to 5 MHz, enhancements in transmission rates, the rake diversity effect (Section 6.4), and the statistical multiplexing effect (Section 9.1.3) are accomplished. A synchronized or nonsynchronized system between base stations can be selected. The nonsynchronized system enhances freedom such as in cell planning and eases a flexible deployment of a cellular system. Furthermore, coherent demodulation with a dedicated pilot signal in time domain intending introduction of an array antenna system and the base station transmit diversity to achieve highly reliable signal transmission without increasing the complexity of an MS are adopted.

TABLE 9.15 IMT-2000 system

Recommended system name	Proposed system name	Multiple access
IMT-DS	W-CDMA	CDMA
	UTRA-FDD	
IMT-MC	cdma2000	CDMA
IMT-TC	UTRA-TDD	CDMA
	TD-SCDMA	TDMA
IMT-SC	UWC-136 (EDGE)	TDMA
IMT-FT	DECT	TDMA
		FDMA

**TABLE 9.16 Major specifications for W-CDMA system**

Frequency band	1920–1980 MHz (up), 2110–2170 MHz (down)
Access scheme	DS-CDMA
Duplexing scheme	FDD
Spreading bandwidth	5 MHz
Chip rate	3.84 M chips/s
Carrier spacing	200 kHz
Data speed	2M Maximum bits/s (maximum)
Frame length	10, 20, 40, 80 ms
Data error correcting codes	Turbo code, convolutional code
Data modulation scheme	BPSK (up), QPSK (down)
Spreading modulation	HPSK (up), QPSK (down)
Inter base station synchronization	Non-coherent and coherent modes are selectable
Voice coding scheme	AMR (1.95 k–12.2 k bits/s)

BPSK in the uplink and QPSK in the downlink are used for data modulation. For spectrum spreading, H(hybrid)PSK in the uplink and QPSK in the downlink with a chip rate of 3.84 M chips/s are adopted, where the HPSK is expressed as follows [71]:

$$c_{\text{long}}(i) = C_{\text{long},1}(i) (1 + j(-1)^i c_{\text{long},2}(2 \lfloor i/2 \rfloor)) \quad (i = 0, 1, 2, 3, \dots)$$

where  $C_{\text{long},1}(i)$  and  $C_{\text{long},2}(i)$  are random (spreading) sequences that take  $\pm 1$ . The symbol denotes  $\lfloor i \rfloor$  an integer not larger than  $i$ . The HPSK is a modulation scheme, which selects QPSK or  $\pi/2$  shifted QPSK in turn at an even and an odd time instants. Thus, the probabilities that its signal trajectory passes through the origin become one half and therefore the peak-to-average power ratio decreases and the power amplification in a mobile terminal becomes easier. The chip pulse waveform is the impulse response of a filter with square root of a raised cosine roll-off characteristics (Eq. 8.5). Turbo (convolutional) codes are used for the error correction scheme. Voice coding system is adaptive multirate (AMR; Section 7.7.6) with adaptive coding rates that changes from 1.95 to 12.2 kbps.

IMT-MC is an expansion of cdmaOne (IS-95) by using three carriers simultaneously and is called also as CDMA2000. The subset of CDMA2000 with  $N(=1, 3)$  subcarrier is called CDMA2000 Nx RTT (RTT: radio transmission technologies). Usually, notation CDMA2000 Nx is used. Actually, a system with  $N=1$  was brought into commercial use. CDMA2000 1X enhanced the maximum transmission rate in both downlink and uplink to 153 kbps while keeping the backward compatibilities with cdmaOne (IS-95).

IMT-TC (time code) includes UTRA-TDD and TS-CDMA, which are proposed by Europe and China, respectively. Both of them are time-division duplexing (TDD) systems, where transmit and receive signals are multiplexed in a time-division manner. A TDD system has advantages that a transmit and receive diversity system is realized using multiple antennas only at base stations and that up- and downlink



traffic imbalance is easily coped with (Section 7.1). The multiple access is a TDMA and CDMA combined system, where a signal frame consists of multiple time slots, and each of the slots is further code division multiplexed.

IMT-SC is the evolutionary system from UWC-136, which is the evolutionary system of IS-136, and from the EDGE evolved from GSM to meet the 3G system requirements. IMT-frequency time (FT) is the 3G version of the European DECT system (Section 9.6.2).

License allocations of 3G systems were desired by many service operators, reflecting the success in 2G system business and the so-called IT bubbles. The prices for the licenses were increased in Europe due to introduction of auction system for the licenses. The commercial deployments of 3G systems were behind the scheduled year 2000. One reason for this is due to the collapse of the IT bubbles. In Japan, NTT Docomo introduced a commercial system under the name of FOMA firstly in the world. The penetration speed was slow at the beginning due to insufficient maturity of terminals. Later, an operator AU introduced a cdma2000 system at the frequency band for the 2G systems. In 2004, the 3G systems went into a full-blown penetration with heated competence between FOMA and cdma2000.

As the frequency bands for IMT-2000, an 800 MHz band (806–960 MHz), a 1.7 GHz band (1710–1885 MHz), and a 2 GHz band (2500–2690 MHz) were added at WRC-2000 (world radio communication conference) held from March to June 2000.

### 9.7.6 Evolution of 3G Systems

Of standardized 3G systems, the major systems are W-CDMA proposed by the 3GPP and cdma2000 by 3GPP2. To achieve a higher packet data transmission, these systems are expanded as mentioned in the following [12, 13].

The cdma2000 took the initiative. The evolutionary system from cdma2000 1X is called CDMA2000 1X EV, which includes CDMA2000 1X EV-DO (data only or data optimized) and CDMA2000 1X EV-DV (data and voice). The former uses a carrier that is specialized for data transmission, and the latter transmits data and voice simultaneously on a carrier. CDMA2000 1X EV-DO is based on the HDR system developed by Qualcomm Inc. [46, 47]. The system transmits high-speed data and low-speed voice signals exclusively on separate dedicated carriers. The reason for this is that the circuit efficiency is enhanced with the separate transmission since voice and data signals require different conditions such as transmission rates, delay, and error rates.

In an HDR down channel, full power transmission is done for each terminal without transmit power control. The signals are multiplexed in time domain instead of code division multiplexing. Furthermore, modulation schemes (QPSK, 8PSK, and 16QAM), error-correcting coding rates, and packet length are changed adaptively to circuit qualities. The maximum transmission rate becomes 2.5 Mbps for a high-quality channel, and the minimum rate is 32 kbps for a low-quality channel. The channel qualities are informed to a base station from terminals by measuring the received pilot signal transmitted from the base station.

Scheduling to give a transmission chance for terminals with different channel qualities becomes an important issue to design a system. Matters to be considered in the scheduling are the average throughput as a whole system and fairness in the channel assignment. If we make emphasis on the throughput, terminals with a high-quality channel are assigned with a high priority. However, terminals with a low-quality channel are scarcely assigned a transmission chance, and the latency, time to complete transmission, becomes long, and consequently the fairness becomes worse. The HDR system adopt a scheme, the PFS (Section 9.1.3), where a transmission right is assigned to a terminal that shows the highest value of a criterion that is defined for each terminal as the ratio of a transmission rate given by the channel quality at the present time to an average transmission rate obtained so far. A user terminal with a low average throughput is given a high priority to keep the fair access to the channel. If the average throughputs are the same between different terminals, a user with the highest transmission rate capability is selected, that is, the multiuser diversity (Section 7.1) [51, 52] is accomplished, and therefore, the throughput of a system as a whole is increased. Since a relatively long time delay is admissible in data transmission, ARQ system (Section 7.4.8) is used in HDR. From the same reason, turbo codes with interleaving and iterative decoding are adopted for a high error correcting capability in spite of their long latency.

In W-CDMA system, high-speed downlink packet access (HSDPA) was standardized first. The most important modification is standardization of high-speed downlink shared channel (HS-DSCH), where the channel is shared with users in a time-division manner similar to HDR. The time-division slot length is short as 2 ms to adapt to time variation of the channel. Signal spectrum spreading is done with codes with a code length of 16. The spreading codes #1–#15 are used for a user and the remaining one is used for such as a channel control purpose. Channel allocation (scheduling) to users is carried out adaptively to channel conditions similarly to HDR. As a modulation scheme, 16QAM is added to conventional QPSK, and one of them is selected depending on the channel conditions. The maximum transmission rate in the downlink is 14.4 Mbps. As a countermeasure against transmission errors, hybrid ARQ (Section 7.4.8) is adopted, where forward error correction and retransmission are combined. In case of retransmission, a soft combining is adopted, where the retransmit data are combined with the prior received data as analog signals. The so-called incremental redundancy (IR) system is introduced, where the retransmit data are different from the prior data (Section 7.4.8).

In the uplink, high-speed uplink packet access (HSUPA) is standardized. Similarly to HSDPA, high-speed scheduling and a soft combining hybrid ARQ are used. A transmission channel, enhanced dedicated channel (E-DCH), is fixed. The time-division slot length is 2 ms, the same as of HS-DSCH. The uplink is different from the downlink in that (i) transmit power is low, (ii) orthogonality between different user signals is not kept, and (iii) multiple base stations take part in receiving signals from a terminal (soft hand-off). Reflecting these facts, multiplexing multiple user signals on a time slot, adopting adaptive power control, and using different implementation method for hybrid ARQ are introduced.

Transmission rate becomes 5.76 Mbps. HSDPA and HSUPA together are called as high-speed packet access (HSPA).

The later evolutions of the 3GPP system are diverged into two groups. One is HSPA-Evolution (or HSPA+), which is an extended version of the 3GPP, and the other is LTE (long-term evolution). The HSPA-Evolution has a (backward) compatibility with HSPA based on W-CDMA. On the other hand, LTE is a new system abandoning the compatibility. However, new technologies such as MIMO transmission are commonly adopted in the two systems, since these systems are standardized at almost the same time. For a smooth introduction of HSPA-Evolution and LTE, a wire-line (core) network is standardized as Evolved Packet Core System Architecture Evolution.

The HSPA-Evolution downlink MIMO transmission assumes two antennas at both transmit and receive sides. At a user terminal, implementation of the MIMO system is optional for compatibility with the conventional system. At a base station, antenna weighting method and common pilot signal are designed so as to have the compatibility. The antenna weights for a data stream are the same as those for conventional closed loop transmit diversity, that is,

$$w_{11} = \frac{1}{\sqrt{2}}, \quad w_{21} = \left( \frac{1+j}{2}, \frac{1-j}{2}, \frac{-1+j}{2}, \frac{-1-j}{2} \right)$$

Antenna weights  $w_{12}$ ,  $w_{22}$  for the other data stream are determined to be orthogonal with  $(w_{11}, w_{21})$  as follows:

$$(w_{11}, w_{21}) (w_{12}^*, w_{22}^*)^T = 0 \quad (T: \text{transpose})$$

This weighting is based on the principle of the eigen-beam space-division multiplexed transmission (Section 7.2.3) using the aforementioned approximate eigenvectors. In case of one data stream, the MIMO system becomes a maximum ratio combining transmit and receive diversity system (Section 7.2.1).

The downlink maximum transmission rate with MIMO is from 23.4 to 28 Mbps depending on channel conditions. To enhance transmission rates for a terminal without MIMO capability, or a line-of-sight channel, where the MIMO transmission is not effective (Section 7.2.3), 64QAM and 16QAM are added in the downlink and the uplink, respectively. Then the maximum transmission rates become 21 and 11 Mbps in the downlink and uplink, respectively. The transmission time delay is shorted below 50 msec.

The LTE aims to accomplish a maximized evolution by abandoning compatibility with conventional W-CDMA and HSPA. It assumes (i) all IP signal transmission, (ii) flexible use of frequency bands, and (iii) emphasis on communication quality for terminals at a cell edge. The IP transmission includes data and voice as voice over IP (VoIP) as well. Then a network is implemented without depending on conventional integrated services digital network (ISDN). Frequency bands are prepared as 1.4, 3, 5, 10, and 20 MHz. As transmit/receive duplexing schemes, both frequency-division duplex (FDD) and TDD are standardized. The system requirements are summarized in Table 9.17.

**TABLE 9.17 Required conditions for LTE system**

Maximum data speed	100 Mbps (down) 50 Mbps (up)
Radio transmission delay (one way)	Less than 5 ms
User data speed	
Cell edge	2–3 times (down) 2–3 times (up)
Cell average	3–4 times (down) 2–3 times (up)
Frequency usage efficiency	3–4 times (down) 2–3 times (up)

Data speed and frequency usage efficiency are against HSPA/HSUPA (Rel. 6).

As signal multiplexing methods, MC OFDM (Section 6.5) in the downlink and SC-FDMA (Section 6.6) in the uplink are adopted. The reason why the OFDM was adopted is that effects of intersymbol interference due to high-speed (wideband) transmission are easily suppressed (Section 6.5.1). The adoption of the SC transmission in the uplink aims the suppression of peak-to-average power ratio with an emphasis on the power efficiency of transmit power amplification in a terminal. The intersymbol interference of the signal is equalized with signal processing possibly as frequency domain equalization (Section 7.3.7) at a base station.

The subcarrier spacing is 15 kHz, and 12 continuous subcarriers are used as a resource block resulting a bandwidth of 180 kHz. The central (dc) subcarrier is a carrier without signal modulation. This is intended to make a quadrature detection to easily extract the inphase and quadrature-phase components by using a local signal with the same frequency as the dc subcarrier. The number of resource blocks is variable from 6 to over 100. Then useful signal bandwidth becomes from 1 to 20 MHz.

In time domain, a signal frame with a length of 10 ms is used. A frame consists of 10 subframes with a 1 ms length. A subframe is divided into two slots with a length of 0.5 ms. The frame structure is the same for the up- and downlinks in an FDD system. In a TDD system, the same frame structure is used except that the subframes are assigned for the up- or downlinks.

In a downlink OFDM channel, two modes are prepared, one of which has a slot consisting of seven symbols with a cyclic prefix (CP) with 4.7  $\mu$ s and the other has a slot of six symbols with a 16.7  $\mu$ s CP. A wider delay spread in a channel is coped with the longer CP.

As modulation schemes, QPSK, 16QAM, and 64QAM are used. In a down channel, MIMO multiplex transmission with four parallel data streams per user is possible. An uplink, however, never adopts a multiplex transmission. With these technologies, the maximum transmission rates become 300 Mbps in the downlink and 75 Mbps in the uplink. The frequency usage efficiencies are 15 and 3.75 bps/Hz, respectively, in the downlink and uplink.

Adaptation of scheduling and transmission rates to channel qualities and hybrid ARQ with soft combining are used as the conventional systems. A multicast/broadcast

single frequency network, where multiple base stations transmit the same signal coherently, is also provided enjoying a benefit of OFDM transmission (Section 6.5.1).

The increase of the maximum transmission rates owes majorly to the increased number of transmit/receive antenna elements and modulation levels. User transmission rates (cell average or cell edge) and frequency efficiency are improved with a comprehensive effect of all the techniques.

The evolution of the LTE continued. On fixing the developments targets for LTE-Advanced in ITU-R, the 3GPP started standardization process of the LTE-Advanced, which is an evolutionary system from LTE. The specifications were issued in March 2011 and recommended together with Worldwide Interoperability for Microwave Access (WiMAX) 2 by ITU-R as the IMT-Advanced. The LTE-Advanced has a backward compatibility with LTE. The performance target of the LTE-Advanced is the maximum transmission rate of 1 Gbps and frequency efficiency of 6.75 bps/Hz. The proposed LTE-Advanced achieves 1 Gbps in the downlink and 500 Mbps in the uplink. The corresponding frequency efficiencies are 30 and 15 bps/Hz, respectively. Such a high-speed transmission and a high-frequency efficiency are attained through bandwidth expansion up to 70 MHz and introduction of  $4 \times 4$  MIMO transmission. The bandwidth expansion is carried out by collecting multiple carrier bands of LTE (carrier aggregation).

To improve communication quality at a cell edge, CoMP [53] is used, where multiple base stations cooperatively transmit/receive the same signal to/from a user terminal: In the downlink, base stations in different cells transmit a signal to the terminal at the same time or being switched fast, or channel allocation scheduling or beam forming to control the antenna directive pattern is carried out to prevent from causing interference to terminals belonging to other cells, when transmitting to a terminal at a cell edge. In the uplink, a signal from a terminal is received by multiple base stations (receive site diversity). Deployment of repeaters is effective to improve channel quality of a terminal at a cell edge. The other method is to deploy the so-called picocells, which covers a small but high traffic density area with a low transmit power. To suppress the interference that appears between the picocell and a macrocell, channel allocation scheduling methods become important.

### 9.7.7 WiMAX

The IEEE 802 committee is doing standardization activities of wireless communication systems as well as the 3GPP and 3GPP2 forums. The WiMAX was proposed by the WiMAX forum based on IEEE 802.16 standardization (March 2004). This system assumes wireless services with fixed terminals. The OFDM with subcarrier spacing of 10.94 kHz is used for multiplexing in both downlink and uplink. The bandwidths are 5, 7, 8.75, and 10 MHz. The maximum transmission rate is 70 Mbps in both links. As error correction schemes, convolutional (turbo) code, block turbo code, and LDPC are used. For transmit/receive duplexing, TDD, FDD, and half-duplex FDD are prepared. As multiple antenna systems, beam forming, space-time coding, and MIMO space-division multiplexing are adopted.

The extended system to accommodate moving terminals by adding handover capability to WiMAX is standardized as IEEE 802.16e (or mobile WiMAX) in December 2005. The successor to the mobile WiMAX to fit the IMT-Advanced is IEEE 802.16m or WiMAX 2 (WiMAX Release 2.0) approved in March 2011. The carrier aggregation to cope with a high-speed transmission using more than 20 MHz is introduced. Later, WiMAX 2+ (Release 2.1) is standardized to give compatibility with TD-LTE system.

9.7.8 The Fourth-Generation System

The systems, LTE-Advanced and WiMAX 2, which meet the IMT-Advanced requirements are called 4G system in narrow sense. The 4G system in a wide sense includes LTE, WiMAX and HSPA+.

9.8 WIRELESS LOCAL AREA NETWORK

9.8.1 IEEE 802.11 Series

The Institute of Electrical and Electronics Engineers (IEEE) is the largest academic organization in the world with its headquarter in the United States. The IEEE 802 committee is a subgroup with a formal name of IEEE 802 LAN/MAN Standard Committee. The naming “802” is said owing to its beginning in February 1980. The committee consists of many working groups. In the field of wireless communications, 802.11, 802.15, 802.16, and 802.20 are known. Here, wireless systems standardized at those committee are briefly introduced. The IEEE 802.16 (WiMAX) is already described in Section 9.7.

Wireless local area network (LAN) rapidly widespread after its standardization in 1997 owing to IEEE 802.11 working group. The 802.11 further evolved with starting of working groups, 802.11a to 802.11n. In those system series, specifications on the physical layer are summarized in Table 9.18. Communication coverage distance is several 10 m; the maximum transmission rates are 1, 2, 11, 54, and 100–200 Mbps; and the frequency bands are in 2.4 and 5 GHz. Figure 9.27 shows the signal frame format, MAC protocol data unit (MPDU) of 802.11. It consists of the frame body, the frame check sequence to detect receive data errors, and the header part for media access control (MAC).

TABLE 9.18 802.11 series systems summary

	(Maximum) data rate (Mbps)	Modulation scheme	Frequency band
802.11	1, 2	DSSS, FHSS, IrDA	2.4 GHz, infrared
802.11b	5.5, 11	CCK	2.4 GHz
802.11a	54, 48, 36, 24, 18, 12, 9, 6	OFDM	5 GHz
802.11g	(54)	OFDM	2.4 GHz
802.11n	(142)	MIMO	2.4 GHz

The MAC layer scheme is based on carrier sense multiple access/collision avoidance (CSMA/CA; Section 9.1.2), which is an autonomous distributed controlled system, well used in wire-line LAN systems. The polling from an access point (base station) is used as an optional scheme. At the physical layer, 802.11 uses DSSS and frequency hopping spread spectrum (FHSS) in license-free industry, science, and medical (ISM) bands and infrared communications.

Channel spacing is 25 MHz, and three to four carriers are available simultaneously in a given bandwidth. As the system deployment spreads, interference between access points increases and consequently effective transmission rates will drop.

As modulation schemes in the DSSS system, BPSK and QPSK are used for transmission rates of 1 and 2 Mbps, respectively. Spectrum spreading is done with the Barker code of an 11 chip length for both the modulation methods: the chip rate becomes 11 Mchips/s. Figure 9.28 shows the signal format, PLCP protocol data unit (PPDU), for the DSSS system, where preamble and header for physical layer convergence procedure (PLCP) are added to the MPDU (Fig. 9.28) supplied from the MAC layer. The preamble and header are used to receive the MPDU signal at physical layer: the Sync is used to establish synchronization between the transmit and receive sides. Signal frame delimiter (SFD) shows the beginning position of the PLCP header, the signal shows information such as modulation methods, and the Service is prepared for user's use. The length means the MPDU signal length, and the CRC is used to detect if errors exist in the header. The preamble and the header are transmitted with 1 Mbps DBPSK and the MPDU part with 1 Mbps DBPSK or 2 Mbps DQPSK.

To increase transmission rates of 802.11, the 802.11b appeared first (standardized in 1999). This system has two signal formats as shown in Figure 9.29. The preamble and header have the same signal format as the 802.11. The signal format (a) with a long preamble keeps compatibility with 802.11 by using the same modulation scheme for the data part. The signal format (b) with a short preamble shortens the Synch signal and uses 2Mbps DQPSK in the PLCP header to reduce transmit time lengths of these parts. To enhance the transmission rates, complementary code keying (CCK) is adopted as a modulation scheme. The CCK selects a spreading code prepared beforehand according to input data to transmit the data (Fig. 9.30).

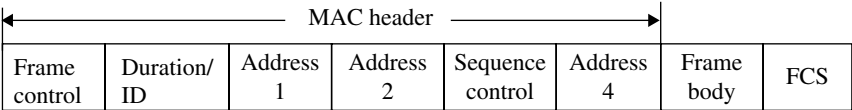


FIGURE 9.27 Frame format for IEEE 802.11 system.

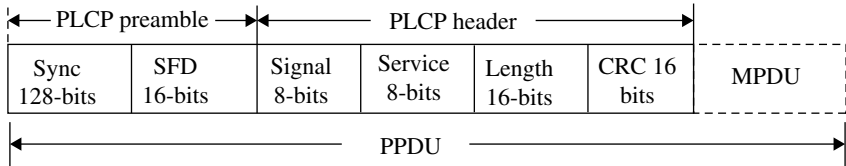
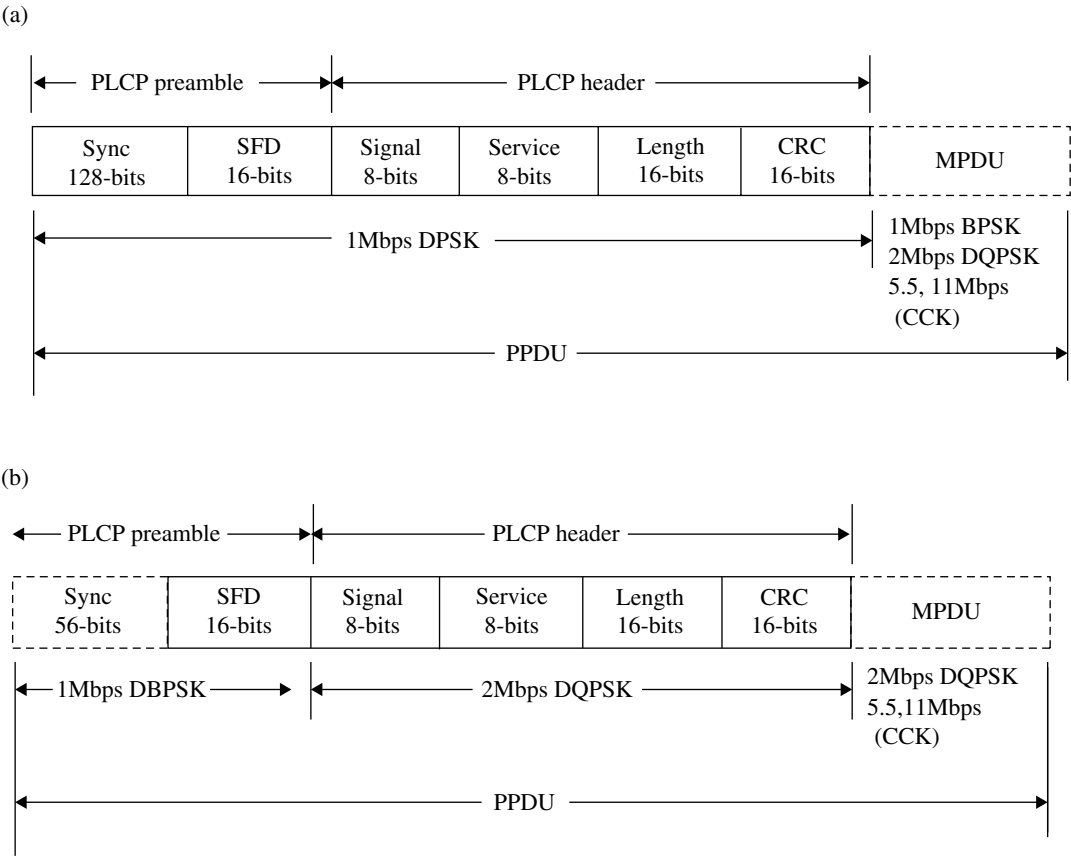
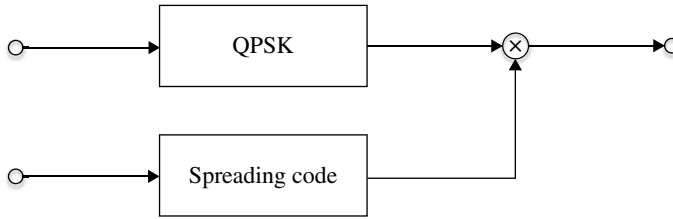


FIGURE 9.28 Frame format for IEEE 802.11 DSSS system.



**FIGURE 9.29** Signal format for IEEE 802.11b DSSS system. (a) Long preamble and (b) short preamble.





**FIGURE 9.30** CCK modulation scheme for 802.11b system (8-chip CCK with chip frequency of 1.735 MHz).

If we use  $2^n$  spreading codes for a modulation with  $m$  bits/symbol, we can express  $n$  bits with the spreading code only and can increase the transmission rate by  $(1 + n/m)$  times for the same chip rate and bandwidth. However, we must pay a penalty of the increase of signal processing complexity due to identifying a spreading code used.

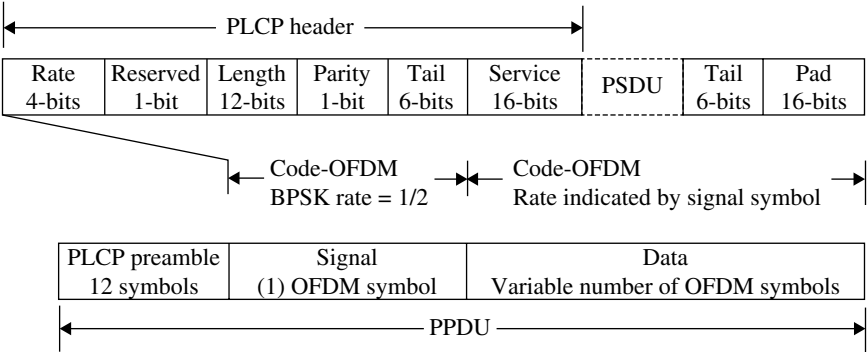
The 802.11b uses a complementary code set with eight complex chips as

$$c = \{e^{j(\phi_1 + \phi_2 + \phi_3 + \phi_4)}, e^{j(\phi_1 + \phi_3 + \phi_4)}, e^{j(\phi_1 + \phi_2 + \phi_4)}, -e^{j(\phi_1 + \phi_4)}, e^{j(\phi_1 + \phi_2 + \phi_3)}, e^{j(\phi_1 + \phi_3)}, -e^{j(\phi_1 + \phi_2)}, e^{j\phi_1}\}$$

By choosing a symbol frequency of 1.375 MHz, the chip rate becomes 11 Mbps as the same as 802.11. Therefore, we can use RF/IF analog circuits commonly with 802.11. Each of the phase terms  $\phi_1 \sim \phi_4$  has 2 bits information and expresses 8 bits data as a whole.  $\phi_1$  never bears information and gives the same phase rotation on a symbol for differential encoding (Section 3.2.7). Since  $\phi_2 \sim \phi_4$  bear data, 6 bits data are expressed with 64 spreading codes. Using a spreading code and QPSK that conveys 2 bits data, the transmission rate becomes 11 Mbps. For a mode of 5.5 Mbps transmission rate, 2 bits are assigned for  $\phi_2 \sim \phi_4$ .

The second evolutionary system of 802.11 is called 802.11a, which was standardized in 1999. A property of the system is that a 5 GHz unlicensed band is assigned: in the United States, 5.15–5.35 and 5.725–5.825 GHz bands called Unlicensed National Information Infrastructure (U-NII) band are used. The other property is that OFDM (Section 6.5.1) is adopted as modulation scheme. OFDM is an MC parallel signal transmission system with multiple carriers. As a common advantage for MC systems, effects of interference in a time-dispersive multipath channel are mitigated: in the standardized OFDM system with a rectangular pulse waveform and insertion of a guard time between each symbol, the receiver never suffers from intersymbol interference as long as the maximum time delay difference in the multipath channel stays within the guard time interval. Actually, a copy of the last part of symbol signal is inserted into the guard time to avoid the discontinuity problem for a receiver with the DFT signal processing.

The signal format of 802.11a is shown in Figure 9.31. The signal frame consists of PLCP preamble and PLCP header and data as well as 802.11. However, the contents are different, and no compatibility exists between them. The time length of the preamble is 16 ms and is divided with an equal time length into a



**FIGURE 9.31** Signal format for IEEE 802.11a system.

short-symbol part and a long-symbol part. The short-symbol part is used for automatic gain control (AGC) and a coarse estimation of channel characteristics. A fine estimation of the channel characteristics is carried out using the long-symbol part and the results of the AGC and the coarse estimation. The short- and long-symbol parts consist of 12 and 53 subcarriers, respectively. A 6 Mbps BPSK modulation is used for both parts.

The signal part in the header includes 24 bits information of transmission rates and length of the data (PSDU). In the signal and data parts, signals encoded with convolutional coding are sent with 48 data subcarriers and 4 pilot subcarriers. In the signal part, coding rate is 1/2, modulation is BPSK, and transmission rate is 6 Mbps. In the data part, coding rates are 1/2, 2/3, and 3/4; modulation schemes are BPSK, QPSK, 16QAM, and 64QAM; and the transmission rates become 6, 9, 12, 18, 24, 36, 48, and 54 Mbps according to a combination of the coding rates and modulation schemes.

On spreading of IEEE 802.11b, 802.11g was standardized in 2003. The frequency band is in 2.4 GHz, the same as 802.11b. As the modulation schemes, CCK from 802.11b and OFDM from 802.11a are mandatory. As the transmission rates, 1, 2, 5.5, and 11 Mbps in 11b mode and 12 and 24 Mbps in 11a mode are mandatory, and optional are 9, 18, 36, 48, and 54 Mbps. The 802.11g has a compatibility with 11b.

To further enhance transmission rates, 802.11n was standardized in September 2009. It aims the maximum and effective rates of 600 and 100 Mbps, respectively. Frequency bands are in 2.4 and 5 GHz. This system has compatibilities with 802.11a, 11b, and 11g at the physical and MAC layers. The approaches to the higher transmission rates are based on channel bonding where multiple bands are used simultaneously (up to 40 MHz), use of higher-level modulation schemes, and space-division parallel stream (up to 4) transmission on the same frequency with MIMO (4×4) systems.

802.11ac is the successor to 802.11n. It aims the maximum rate of 6.93 Gbps. The frequency band is 5 GHz. The maximum bandwidth is expanded to 80 or 160 MHz from 40 MHz. As a modulation scheme, 256QAM is added. The number of MIMO antennas is increased to 8 from 4. The multiuser MIMO system (Section 7.2.3) is introduced.

802.11ad aims also a high transmission rate as the maximum rate of 6.8 Gbps. The frequency band is 60 GHz, where a wide bandwidth of 9 GHz is allocated in an

unlicensed band: this is in contrast to the bandwidth of 0.4GHz for a 5GHz system as 802.11ac. Except for the frequency band, the 802.11ad is based on the same technologies as 802.11ac.

9.8.2 Bluetooth

IEEE 802.15 series deals with the so-called personal area network (PAN), of which communication coverage distance is around 10 m. IEEE 802.15.1 defines a system called the Bluetooth. The developments of the system had been initialized in 1998 by a consortium of five companies, that is, Ericsson, Nokia, Motorola, Intel, and Toshiba. Later, it was handed to 802.15.1, and version 1.1 was standardized in 2002. The specifications together with those for later versions are shown in Table 9.19. Its use was spread for such as a hand-free system, which makes a cell phone operation without using a hand(s) possible with a microphone and a headphone, a measure to give a cordless telephone function to a cell phone, and implementations to PC for such as a wireless mouse. Modulation methods are defined for the basic rate (mandatory) and the enhanced data rate (EDR, optional). The symbol rate is 1 M symbol/s for both methods. For the basic rate, Gaussian-filtered FSK (GFSK; Section 6.2.2) is used, which is a power-efficient constant envelope modulation. For the EDR,  $\pi/4$  shifted DQPSK and 8DPSK are used for transmission rates of 2 and 3 Mbps, respectively. To suppress interference to/from other systems, FH technique is adopted, where an adaptive selection of FH frequencies is also carried out not to use those with high interference.

9.8.3 UWB

The ultra-wideband (UWB) system is originally an impulse radio, where wideband, and therefore narrow, pulse signals are sent directly without modulation. It was developed for military use as a radar to detect persons or objects behind a wall by using its high time resolution capability. It has an advantage that interference to other systems is kept low owing its low power spectrum densities due to the wide bandwidth. The other advantage is free from modulation/demodulation systems. IEEE 802.15.3a discussed two candidates as the UWB as shown in Table 9.20. However, the WG failed to accomplish standardization and disbanded in 2006.

TABLE 9.19 Bluetooth system major specifications

Frequency band	2.4 GHz ISM band
Transmit power	1 mW (radius 10 m) to 100 mW (radius 100 m)
Modulation scheme	GFSK/FHSS (first/second stage) Hopping speed:1600 hops/s)
Maximum data transfer rate	1 Mbps (version 1.1/1.2) 2Mbps or 3 Mbps (version 2.0+EDR)
Channel spacing	1 MHz
Number of simultaneously transmit terminal	8 terminal/ch (32ch)

9.8.4 ZigBee

IEEE 802.15.4 standardized the ZigBee in 2005 for application to automation at homes and factories (HA and FA). Radio specifications are summarized in Table 9.21. The ZigBee can be applied to a control system through a wireless sensor network enjoying its benefits of low cost and low power consumption.

9.8.5 BWN

IEEE 802.16 series deal with the so-called metropolitan area network (MAN) with a wide service area. The other naming was broad wireless access (BWN). Table 9.22 summarized the specifications for this series standardization. The 802.16 assumes a line-of-sight communication channel. As modulation scheme, SC system is used. 802.16a assumes a non-line-of-sight communication channel and uses SC, OFDM, and OFDMA systems to cope with dispersive multipath propagation. The 802.16-2004 is called fixed WiMAX, which is a merged system of 802.16 and 802.16a. The 802.16e is called mobile WiMAX, which copes with mobile communication environments by introducing such as handover procedure. The system is explained in Section 9.7.7.

TABLE 9.20 Two UWB systems discussed in IEEE 802.15.3a

	DS-UWB system (direct spreading UWB)	MB-OFDM system (multiband OFDM)
Frequency	3.1–5.15 GHz (low band) 5.825–10.6GHz (high band)	3.168–5.820 GHz (low band) 5.544–10.296 GHz (high band)
Number of channels	2	13
Occupied bandwidth	1.368 GHz (low band) 2.736 GHz (high band)	528 MHz
Transmission speed	25–450 bps (low band) 25–900 bps (high band) ~1350 Mbps (multiband)	55–480 Mbps
Wireless access scheme	TDMA-TDD	
Modulation scheme	DS-SS (BPSK, QPSK, M-BOK)	MB-OFDM (BPSK, QPSK)

TABLE 9.21 ZigBee system specifications

Frequency	2.4 GHz (global)	915 MHz (USA)	868 MHz (Europe)
Number of channels	16	10	1
Modulation scheme	O-QPSK	BPSK	BPSK
Spreading scheme	DS-SS	DS-SS	DS-SS
Chip rate(kchip/s)	2000	600	300
Symbol rate (sps)	62.5	40	20
Data rate(kbps)	250	40	20
Transmit power		> −3 dBm	

**TABLE 9.22 Major specifications for IEEE 802.16 series systems**

Frequency band	802.16 10–66 GHz	802.16a 2–11 GHz (including un-licensed band)	802.16–2004 ~11 GHz (including un-licensed band)	802.16e ~6 GHz
Service situations	Fixed	Fixed	Nomadic, fixed	Mobile, portable, nomadic, fixed
Propagation environments	LOS	NLOS	NLOS	NLOS
Channel bandwidth	20, 25, 28 MHz	1.25–20 MHz	1.25–20 MHz, 25, 28 MHz	1.25–20 MHz
Physical layer scheme	SC	SC, OFDM, OFDMA	SC, OFDM, OFDMA	SC, OFDM, OFDMA
Maximum transmission speed	135 Mbps (for 28 MHz bandwidth)		75 Mbps (for 20 MHz bandwidth)	
Specified date	Dec. 2001	March 2003	June 2004	Dec. 2005

LOS, line of sight; NLOS, non-line of sight.

**TABLE 9.23 IEEE 802.20 required specifications**

Maximum moving speed	250 km/h
Frequency usage efficiency	More than 1 bit/s/Hz/cell
Maximum user speed	>1 Mbps (down) >300 kbps (up)
Maximum transmission speed per cell	>4 Mbps (down) >800 kbps (up)
MAC frame round trip time	<10 ms
Highest commercial frequency	<3.5 GHz
Duplex system	FDD or TDD
Frequency assignment	Licensed frequency band

### 9.8.6 MBWA

IEEE 802.20 started in 2002 aiming at the mobile broadband wireless access (MBWA), which gives high-speed and high-frequency efficiency services in fast-moving and wide area environments [72]. The requirements for the systems are issued in 2004 as shown in Table 9.23. After a short time suspension, it was standardized in August 2008. The standard has two modes of operation. One is a wideband mode and the other is a 625k-MC mode. There is no compatibility between the two modes. The wideband mode is an OFDMA system with FH. It is designed to operate for all FDD and TDD bandwidths. The bandwidths are changeable between 5 and 20 MHz with a granularity of 1.25 MHz. The 625k-MC mode is designed with 625 kHz carrier bandwidth supporting aggregation of multiple carriers. It is operated only in TDD. Adaptive array antenna systems are used for high spectrum efficiency and high mobility.

## APPENDIX 9.A POISSON ARRIVAL RATES

The occurrence rates for a random event can be calculated as follows. Consider a time length of  $t$  and divide it into  $n$  short periods of  $\Delta t$  ( $t = n\Delta t$ ) where the time period is short enough such that only one event can occur in a period. Let us assume the occurrence rates per unit time as  $\lambda$ . Then the probability that an event occurs during a short time period  $\Delta t$  becomes  $\lambda\Delta t$ . Since events occur independently, the probability that  $k$  events occur in the time length  $t$  is expressed as

$$v_k(t) = \lim_{\Delta t \rightarrow 0} \binom{n}{k} (\lambda\Delta t)^k (1 - \lambda\Delta t)^{n-k} = \frac{(\lambda t)^k}{k!} \lim_{n \rightarrow \infty} \left(1 - \lambda \frac{t}{n}\right)^n$$

where

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

Using  $\lim_{n \rightarrow \infty} (1 + x/n)^n = e^x$ , we have

$$v_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

This stochastic process is called the Poisson arrival process.

## REFERENCES

- [1] Jakes WC, editor. *Microwave Mobile Communications*. New York: Wiley; 1974.
- [2] Balston DM, Macario RCV, editors. *Cellular Radio Systems*. Boston: Artech House; 1993.
- [3] Lee WCY. *Mobile Cellular Telecommunications*. 2nd ed. New York: McGraw-Hill; 1995.
- [4] Lee WCY. *Mobile Communications Design Fundamentals*. 2nd ed. New York: Wiley; 1993.
- [5] Stüber GL. *Principles of Mobile Communication*. Boston: Kluwer Academic Publishers; 1996.
- [6] Rappaport TS. *Wireless Communications, Principles and Practice*. Englewood Cliffs: Prentice Hall; 1996.
- [7] Steal R, editor. *Mobile Radio Communications*. London: Pentech Press Publishers; 1992.
- [8] Yacoub MD. *Foundations of Mobile Radio Engineering*. Boca Raton: CRC Press; 1993.
- [9] Hess GC. *Land-Mobile Radio System Engineering*. Boston: Artech House; 1993.
- [10] Pahlavan K, Levesque AH. *Wireless Information Networks*. New York: Wiley; 1995.
- [11] Gibson JD, editor. *The Mobile Communications Handbook*. Boca Raton: CRC Press; 1996.
- [12] Dahlman E, Parkvall S, Skold L. *LTE/LTE-Advanced for Mobile Broadband*. Oxford: Academic Press; 2011.
- [13] Dahlman E, Parkvall S, Sköld J, Beming P. *3G Evolution, HSPA and LTE for Mobile Broad*. Oxford: Academic Press; 2007.
- [14] Nagata Y, Akaiwa Y. Analysis for spectrum efficiency in single cell trunked and cellular mobile radio. *IEEE Trans Veh Technol* 1987;VT-36:100–113.
- [15] Stjernvall JE. Calculation of capacity and cochannel interference in a cellular system. *Proceedings of the First Nordic Seminar on Digital Land Mobile Radio Communication*; Espoo, Finland, 1985. p. 209–217.
- [16] Kanai T. Channel assignment for sector cell layout. *Trans IEICE* 1990;J73-B-II:595–601.
- [17] Yamada Y, Ebine Y, Tsunekawa K. Base and mobile station antennas for land mobile radio systems. *Trans IEICE* 1991;E74:1547–1555.
- [18] Tong F, Akaiwa Y. Effect of beam tilting on bit-rate selection in mobile multipath channel. *Proceedings of the 3rd International Conference on Universal Personal Communications*; September–October 1994; p. 225–229.
- [19] Halpern SW. Reuse partitioning in cellular systems. *Proceedings of the IEEE Vehicular Technology Conference*; May 1983; 322–327.
- [20] Mikulski JJ. Dyna T\*A\*C cellular portable telephone system experience in the U.S. and the UK. *IEEE Commun Mag* 1986;24:40–46.
- [21] Lathin L. Radio network structures for high traffic density areas. *Proceedings of the Third Nordic Seminar on Digital Land Mobile Radio Communication*, No. 14.10; 1988.
- [22] Worsham J, Avery J. A cellular band personal communications systems. *Proceedings of the Second International Conference on Universal Personal Communications*; 1993. p. 254–257.

- [23] Kinoshita Y, Tsuchiya T, Ohnuki S. Common air interface between wide-area cordless telephone and urban cellular radio: frequency channel doubly reused cellular systems. *Trans IEICE* 1993;J76-B-II:487–495.
- [24] Furukawa H, Akaiwa Y. A microcell overlaid with umbrella cell system. *Proceedings of the IEEE Vehicular Technology Conference*; June 1994. p. 1455–1459.
- [25] Nakajima N. Evolution of cell layout techniques-toward microcell systems. *NTT DoCoMo Tech J* 1993;1:21–29.
- [26] Taketsugu M, Ohteru Y. Holonic location registration/paging procedure in microcellular systems. *IEICE Trans Fundam Electron Commun Comput Sci* 1992;E75-A:1652–1659.
- [27] Wang JZ. A fully distributed location registration strategy for universal personal communication systems. *IEEE J Selected Areas Commun* 1993;11:850–860.
- [28] Tanenbaum AS. *Computer Networks*. Englewood Cliffs: Prentice-Hall; 1981.
- [29] Hammond JL, O'Reilly PJ. *Performance Analysis of Local Computer Networks*. Reading: Addison-Wesley; 1986.
- [30] Akaiwa Y, Furuya Y, Kobayashi K. Method of determining optimal transmission channel in multi-station communications system. US patent 4,747,101, 1988.
- [31] Furuya Y, Akaiwa Y. Channel segregation, a distributed adaptive channel allocation scheme for mobile communication systems. *Proceedings of the second Nordic Seminar on Digital Land Mobile Radio Communication*, Stockholm, October 1986; also in. *IEICE Trans* 1991;E-74:1531–1537.
- [32] Akaiwa Y, Andoh H. Channel Segregation—a self-organized dynamic channel allocation method: application to TDMA/FDMA microcellular system. *IEEE J Selected Areas Commun* 1993;11:949–954.
- [33] Akaiwa Y, Furukawa H. Application of channel segregation for automatic channel selection free from intermodulation interference. *Proceedings of the Seventh IEEE International Symposium on Personal, Indoor and Mobile Radio Communication*; October 1996. p. 1235–1236.
- [34] Kanai T. Autonomous reuse partitioning in cellular systems. *Proceedings of the IEEE Vehicular Technology Conference*; May 1992. p. 782–785.
- [35] Furukawa H, Akaiwa Y. Self-organized reuse partitioning a dynamic channel assignment method in cellular system. *Proceedings of the IEEE Vehicular Technology Conference*; May 1993. p. 524–527.
- [36] Kataoka N, Miyabe M, Fujino T. A distributed dynamic channel assignment scheme using information of received signal level. *IEICE Technical Report RCS-93-70*; November 1993. p. 1–7.
- [37] Furukawa H, Akaiwa Y. Self-organized reuse partitioning (SORP), a distributed dynamic channel assignment method. *Technical Report of IEICE, RCS-92-126*; January 1993. p. 61–66.
- [38] Goodman DJ, Valenzuela RA, Gayliard KT, Ramamurthi B. Packet reservation multiple access for local wireless communications. *IEEE Trans Commun* 1989;COM-37:885–890.
- [39] Feher K, editor. *Advanced Digital Communications*. Englewood Cliffs: Prentice-Hall; 1987.
- [40] Frullone M, Riva G, Grazioso P, Crciofi C. Self-adaptive channel allocation strategies in cellular environments with PRMA. *Proceedings of the IEEE Vehicular Technology Conference*; June 1994. p. 819–823.



- [41] Beck R, Panzer H. Strategies for handover and dynamic channel allocation in micro-cellular mobile radio systems. Proceedings of the IEEE Vehicular Technology Conference; May 1989. p. 178–185.
- [42] Yokayama M. Decentralization and distribution in network control of mobile radio communications. Trans IEICE 1990;E73:1579–1586.
- [43] Katzela I, Naghshineh M. Channel assignment schemes for cellular mobile telecommunication systems: a comprehensive survey. IEEE Personal Communications; June 1996. p. 10–31.
- [44] Amitay N. Distributed switching and control with fast resource assignment/handoff for personal communications systems. IEEE J Selected Areas Commun 1993;11:842–849.
- [45] Chung JC-I. Performance issues and algorithms for dynamic channel assignment. IEEE J Selected Areas Commun 1993;11:955–963.
- [46] Bender P, Black P, Grob M, Padovani R, Sindhushyana N, Viterbi A. CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users. IEEE Commun Mag 2000;38:70–77.
- [47] Jalali A, Padovani R, Pankaj R. Data throughput of CDMA–HDR a high efficiency-high data rate personal communication wireless system. Proceedings of the IEEE Vehicular Technology Conference; 2000. p. 1854–1858.
- [48] Kushner HJ, Whiting PA. Convergence of proportional-fair sharing algorithms under general conditions. IEEE Trans Wirel Commun 2004;3 (4):1250–1259.
- [49] Kelly F. Charging and rate control for elastic traffic. Eur Trans Telecommun 1997;8:33–37.
- [50] Lau VKN. Proportional fair space–time scheduling for wireless communications. IEEE Trans Commun 2005;53 (8):1353–1360.
- [51] Viswanath P, Tse DNC, Laroia R. Opportunistic beamforming using dumb antennas. IEEE Trans Inf Theory 2002;48 (6):1277–1294.
- [52] Knopp R, Humblet PA. Information capacity and power control in single-cell multiuser communications. Proceedings of the IEEE ICC 1995. p. 331–335.
- [53] Sawahashi M, Kishiyama Y, Morimoto A, Nishikawa D, Tanno M. Coordinated multipoint transmission/reception techniques for LTE advanced. IEEE Wirel Commun 2010;17:26–34.
- [54] 3GPP TR 29.996 V6.1.0 (2003–2009).
- [55] Kuwabara M, editor. *Car Telephone*. Tokyo: IEICE; 1985 (in Japanese).
- [56] Yasuda Y, editor. *Mobile Communications in ISDN Era*. Tokyo: Ohm; 1992 (in Japanese).
- [57] Alshamali A, Macario R. Technical features of the planned European radio messaging system-ERMES. IEEE Vehicular Technology Society News; August 1994. p. 22–25.
- [58] Nakanishi T, Murata E, Honma K, Rikou Y. Digital voice processing land mobile radio. Proceedings of the 1980 International Conference on Security through Science and Engineering; September 1980. p. 58–62.
- [59] Ikoma M, Kimura K, Saegusa N, Akaiwa Y, Takase I. Narrow-band digital mobile radio equipment. Proceedings of the IEEE International Conference on Communication; 1981. p. 23.3.1–23.3.5.
- [60] Hiyama T, Yotsutani A, Kage K, Ichihara M. 4-level FM digital mobile radio equipment. NEC Research and Development, No. 71; October 1983. p. 20–26.

- [61] Nakajima M, Watanabe T, Saka S, Nogami H, Karasawa T. A narrow band digital portable radio using 8 kbps speech CODEC. Proceedings of the Fall National Convention of IEICE, No. B-273; 1990.
- [62] Lycksell E. MOBITEK, a new radio communication system for dispatch traffic. TELE 1983;35 (1):68–75.
- [63] Khan M, Kilpatrick J. MOBITEK and mobile data standards. IEEE Commun Mag 1995;33 (3):96–101.
- [64] Miyamoto T, Tatsumi H, Orikasa H. Tele-terminal system (mobile data communications). Proceedings of the First International Workshop on Mobile Multimedia Communications; December 1993. p. A.1.5-1–A.1.5-8.
- [65] Harris T. Data services over cellular radio. In: Balston DM, Macario RCV, editors. *Cellular Radio Systems*. Boston: Artech House; 1993, Chapter 12.
- [66] Muller NJ. *Wireless Data Networking*. Boston: Artech House; 1995.
- [67] Kinoshita K, Hata M, Hirade K. Digital mobile telephone system using TD/FDMA scheme. IEEE Trans Veh Technol 1982;VT-31:153–157.
- [68] Akaiwa Y, Nagata Y. A linear modulation scheme for spectrum efficient digital mobile telephone systems. Proceedings of the International Conference on Digital Land Mobile Radio Communications, Venice; 1987. p. 218–226.
- [69] Gilhousen KS, Jacobs IM, Padovani R, Viterbi AJ, Weaver LA Jr, Wheatley CE III. On the capacity of a cellular CDMA system. IEEE Trans Veh Technol 1991;VT-40:303–312.
- [70] Nakajima N. Japanese Digital Cellular Radio in: Balston DM, Macario RCV, editors. *Cellular Radio Systems*. Boston: Artech House; 1993; Chapter 10.
- [71] Laird K, Whinnett N, Buljore S. A peak-to-average power reduction method for third generation CDMA reverse links. Proceedings of the of IEEE 49th Vehicular Technology Conference; 1999; p. 551–555.
- [72] Bolton W, Xiao Y, Guizani M. IEEE 802.20: Mobile Broadband Wireless Access. IEEE Wireless Communications 2007;84–95.

# INDEX

- a priori probabilities, 152–6, 232, 419, 432, 443, 446–51
- A-b-S (analysis-by-synthesis), 476
- ACPR *see* adjacent channel leakage power ratio (ACPR)
- Adachi, F., 367
- adaptive
  - array antenna, 275, 457–9, 463, 608
  - automatic equalizer, 355, 401–22
  - cooperation, 569, 573, 574, 576
  - delta modulation, 7, 296, 471, 472, 482, 579
  - differential pulse code modulation, 114, 472–3
  - interference cancellation, 355, 456–69
  - interference suppression, 466–9
  - predictive coding, 473–5
  - predistorter amplifier, 514
  - signal processing, 108–26
  - signal separation, 396
- adaptive differential pulse code modulation (ADPCM), 114, 115, 472, 473, 475, 479, 482, 483, 581, 582
- adaptive predictive coding (APC), 473–5
- additive white Gaussian noise, 40, 183, 272, 434
- adjacent channel leakage power ratio (ACPR), 514
- ADPCM *see* adaptive differential pulse code modulation (ADPCM)
- advanced encryption standard (AES), 180
- AES *see* advanced encryption standard (AES)
- AF *see* amplified and forwarded (AF)
- AGC *see* automatic gain control
- Akaiwa, Y., 306, 314
- ALC *see* automatic level control
- algebraic CELP (ACELP), 479, 481, 592
- aliasing error, 76, 95, 96, 105, 271, 337
- ALOHA system
  - pure ALOHA, 552, 553
  - slotted ALOHA, 552, 553
- alternating-current circuit theory, 56
- alternative-mark-inversion (AMI) code, 145, 146
- AM-AM conversion, 272
- amplified and forwarded (AF), 567
- amplifier

- amplifier (*cont'd*)
  - carrier (class-AB), 523
  - class-AB, 523, 525, 527
  - class AB power, 508
  - class-B, 523
  - class-C, 523, 525, 527
  - common, 8, 494
  - Doherty, 522–5
  - feedback, 507–10
  - feed-forward, 507, 510, 512
  - LINC, 512
  - linear power, 314, 513, 525, 591
  - negative feedback, 508
  - nonlinearity compensated, 507
  - peak, 523–5
  - predistorter, 513
  - push-pull, 522–4
  - saturated power, 220, 282
- amplitude shift keying (ASK), 220, 245, 265, 360
- AM-PM conversion, 272
- AMPS, 577, 587, 588, 590
- antenna aperture, 189, 214, 493
- APC *see* Adaptive predictive coding (APC)
- ARP *see* Autonomous reuse partitioning (ARP)
- ASK *see* Amplitude Shift Keying (ASK)
- automatic
  - gain control, 118, 251, 314, 495, 527, 528, 604
  - level control, 537, 538
- automatic repeat request (ARQ)
  - Go Back N, 452
  - hybrid ARQ, 567, 596, 598
  - Selective Repeat, 452, 453
  - Stop and Wait, 452
- autonomous reuse partitioning (ARP), 554, 557–8
- AWGN *see* additive white Gaussian noise
  
- back-off, 552, 553
- Bay's mixed rule, 168
- BCJR algorithm, 438, 447
- beam tilting, 546–7
- Berrou, C., 430, 451
- binary phase shift keying (BPSK), 139, 221–5, 228, 233–5, 237–9, 241, 246, 249–51, 265, 267–9, 283, 285, 286, 289, 305, 306, 322, 360, 366, 395, 594, 601, 604, 606
- Black, H.S., 71
- block code, 423–6, 444, 446
- bluetooth, 605
- Boltzmann constant, 38, 212
- BPSK *see* Binary phase shift keying (BPSK)
- branch metric, 173, 430, 443
- broad wireless access (BWN), 606
- brute force method, 170, 174
  
- carrier
  - to noise power ratio, 149, 233, 263
  - phase slip, 248, 287
  - recovery, 224, 233–7, 266, 274, 287–9, 291, 371, 532, 533
- carrier aggregation, 599, 600
- Carson's bandwidth, 226
- Cartesian loop, 508, 509
- Cauchy–Schwartz inequality, 377, 381
- causality, 58–9, 86, 94, 115, 399
- CCDF *see* complementary cumulative distribution function (CCDF)
- CCITT, 349, 428, 453, 470, 472, 581
- CCK *see* complementary code keying (CCK)
- CCPSK, 302–4
- CDM *see* Code division multiplexing (CDM)
- CDMA2000
  - Nx RTT, 594
  - 1X EV, 595
  - 1X EV-DO, 595
  - 1X EV-DV, 595
- cdmaOne, 481, 592, 594
- cell cluster size, 544, 548
- cell edge 5% throughput, 575, 576
- cellular concept, 543–51
- cellular data link control (CDLC) system, 580, 581
- cellular digital packet data (CDPD), 580, 581
- CELP *see* code-excited linear predictive (CELP)
- Center for Radio Systems (RCR), 591
- central limit theorem, 41, 51–5, 193
- channel
  - capacity, 155, 183–5
  - segregation, 554–7
  - selection filter, 495, 496, 564

- channel assignment
  - dynamic, 494, 554, 557, 563
  - fixed, 554
  - hybrid, 554
- check bit, 423, 424, 428, 452
- Chinese remainder theorem, 182, 185–6
- CIR *see* signal to interference ratio (CIR)
- circuit-switched system, 552, 563
- circulator, 501
- class II partial response, 262, 294
- click(s), 253–5, 259, 260
  - noise, 257, 259, 263, 370, 482, 483
- clipping, 514, 521, 522
- code(s)
  - class II partial response, 143, 148, 297
  - concatenated, 430, 432
  - convolutional, 385, 423, 429–31, 433, 438, 444, 446, 453, 584, 590, 594, 604
  - CRC, 452
  - cyclic, 426–9, 579
  - duoquaternary, 147–8
  - error-correcting, 183, 320, 423, 424, 453, 595
  - Gold, 328, 329
  - Gray, 138–9, 241, 318
  - Manchester, 139–41, 282, 535, 578
  - multilevel, 137
  - natural binary, 138
  - space-time block, 385
  - space-time trellis, 385
  - systematic, 424, 427, 430, 453
- code division multiplexing (CDM), 182, 183, 324–5, 328, 521
- code-excited linear predictive (CELP), 477–81
- codebook, 477–81
- coding
  - adaptive predictive, 473–5
  - correlative, 141–8, 305
  - duobinary, 142, 144, 148, 294, 298
  - Gray, 138, 242, 245, 395
  - line, 32, 127, 137–48
  - voice, 1, 7, 114, 296, 355, 469–84, 564, 567, 579, 581, 582, 584, 585, 587–92
- coherence bandwidth, 202–3, 264, 267–9, 326, 401
- combined coding and modulation, 453
- comfort noise, 484
- CoMP *see* coordinated multi-point (CoMP)
- complementary code keying (CCK), 600, 601, 603, 604
- complementary cumulative distribution function (CCDF), 316, 335, 522
- constant envelope
  - modulation, 220–221, 275, 282–314, 525, 527, 605
  - PSK, 228, 304, 305
- continuous phase
  - frequency shift keying, 226, 282, 283
  - FSK, 226, 282, 283
  - modulation, 305
- control channels, 587, 592
- convolution integral, 22, 26, 27, 56, 67, 158, 261, 271, 276, 323, 325, 577
- coordinated multi-point (CoMP), 569, 599
- cordless telephone, 472, 482, 581–3, 605
- correlation
  - auto, 78
  - coefficient, 202
  - cross, 27, 49, 79, 119, 195, 205, 324, 328, 466–7, 571
  - of the envelope, 197–8
  - function, 27, 43, 166, 198, 202, 267, 268
  - receiver, 161, 162, 174, 291, 328, 346
- correlative coding, 141–8, 305
- correlative phase shift keying (CORPSK), 304
- cosmic microwave background radiation, 39
- Costas loop, 234–6, 532, 533
- CP *see* cyclic prefix (CP)
- CPFSK, 226, 282, 283
- CRC *see* cyclic redundancy check (CRC)
- CSMA
  - nonpersistent CSMA, 553
  - 1-persistent CSMA, 553
  - p*-persistent CSMA, 553
- CSMA/CD, 553
- CT-2 (second generation cordless telephone), 581, 582
- cyclic
  - prefix (CP), 331–3, 598
  - redundancy check (CRC), 484, 583, 601
- data encryption standard (DES), 178–80
- dc
  - blocking, 535–8
  - offsets, 139, 535–8

- decision making, 113, 127, 128, 143, 149,
  - 167, 169, 172–5, 241, 341, 404,
  - 454, 567
- decode and forward (DF), 567
- DECT *see* Digital European Cordless Telecommunications (DECT)
- Dekker, C.B., 281, 294
- delay
  - line, 68, 69, 88, 130, 246, 529,
  - 530, 533
  - profile, 200
  - spread, 200–204, 585, 586, 591, 598
- delta function
  - Diriclet integral type, properties of, 11
  - periodic, 15
- delta modulation, 7, 296, 471–2
- demand assignment, 551
- demodulator
  - coherent, 221, 233, 234, 532
  - differential, 532, 533
  - noncoherent, 532
- DES *see* data encryption standard (DES)
- descrambling, 3, 178, 179
- despreading, 323, 324, 326
- detection
  - coherent, 139, 148, 224, 233–5, 266,
  - 283, 288, 291, 292, 298–302, 316,
  - 319, 322, 360, 361, 363–5, 373–5,
  - 395, 456
  - delay, 139, 221, 233, 246, 267, 300
  - differential, 148, 224, 233, 246–51,
  - 265–7, 269, 291–3, 297, 300, 302,
  - 306, 311, 316, 318, 319, 321, 360,
  - 364, 366, 533
  - differentially coherent, 233, 246
  - frequency discriminator, 250–264, 266,
  - 298–300, 302, 303, 305, 306, 318
  - phase comparison, 246
- deviated-frequency-locking scheme, 291
- DF *see* Decode-and-forward (DF)
- DFT *see* discrete Fourier transform (DFT)
- diagonalization, 389, 390, 392, 396,
  - 417, 420
- difference equation, 90, 92–4, 562
- differential
  - encoder, 148, 149
  - encoding, 74, 127, 148, 224, 236, 287,
  - 289, 290, 292, 298, 310, 317, 603
  - equation, 63–6, 92, 94
- digital
  - filter, 94–6, 98, 100, 112–14, 118, 338,
  - 339, 343, 368, 475
  - FM, 226, 231–2, 251, 259, 262, 269,
  - 281, 294–308, 310, 312, 314, 366,
  - 367, 369, 506, 537, 579
  - phase modulation, 305, 584
  - predistoeter, 526
  - speech interpolation, 559
- Dirac, P.A.M., 9, 11–13, 21
- direct conversion, 495–500, 533,
  - 534, 578
- direct sequence (DS) system, 323, 324, 327,
  - 565, 566, 590
- directive pattern, 457, 458, 599
- discrete Fourier transform (DFT), 85,
  - 102–8, 270, 271, 333, 337, 339,
  - 343–5, 535, 603
- discrete-time Fourier transform (DTFT),
  - 79–85
- diversity
  - angle, 356
  - cyclic delay, 356
  - equal gain combining, 356, 358–9
  - frequency, 328, 336, 343, 345, 356
  - implicit, 357, 418, 421
  - macro-(or site), 356
  - maximum ratio combining, 358, 597
  - micro-, 356
  - polarization, 356
  - selection (switching) combining, 356
  - space, 356–8, 375
  - time, 356, 357, 579
  - transmission system, 355–75
- Doherty, W.H., 522
- Doppler frequency, 3, 194, 198, 203,
  - 264, 266, 267, 272, 310, 366,
  - 373, 421
- double conversion method, 499–500
- DPD *see* digital predistoeter
- DPM, 305, 584
- DSI *see* digital speech interpolation
- duobinary, 134, 142–6, 148, 228, 229, 262,
  - 294, 295, 298, 412
- duplexer, 493, 494, 501
- dynamic
  - programming, 170
  - range, 115, 204, 251, 282, 314, 316, 472,
  - 481, 527–9, 590

- eavesdropping, 4, 178
- echo canceller, 401
- EFR *see* enhanced full rate (EFR)
- Eigen
  - beam method, 390–392
  - functions, 57, 65, 166
  - value(s), 57, 166
  - vector, 381–3, 389–92, 394, 396, 416–18, 420, 597
- Eigenvector method, 389–90, 392, 396, 420
- 802.11a, 600, 603–5
- 802.11n, 600, 604
- 802.16-2004, 607
- 802.16e, 600, 606, 607
- 8-level PSK (8PSK), 221, 233, 234, 243, 249, 315, 320, 454–6, 592, 595
- energy
  - per bit, 33, 151, 156, 157, 164, 184, 267, 436
  - per bit to noise power density ratio, 149, 245
  - per symbol, 33, 239, 242, 244, 248
  - spectral density, 26–8
- enhanced data rates for global evolution (EDGE), 320, 592
- evolution, 592, 595
- enhanced full rate (EFR), 481, 592
- enhanced variable bit rate coder (EVRC), 481
- envelope detection, 233, 245–6
- equal gain combining, 356, 358–9, 364–6
- equalization, 72, 74, 112, 113, 282, 326, 334, 337, 399, 405–9, 411, 412, 417–21, 519, 528, 598
- equalizer(s)
  - decision feedback, 145, 356, 409–10, 419, 421, 537, 586
  - frequency domain, 356, 418–20
  - linear, 40, 402–5, 408, 409, 414, 419, 420
  - negative feedback, 403
  - transversal, 403–5
  - turbo, 419
  - Viterbi, 410, 421
  - zero-forcing, 408
- equivalent
  - baseband complex expression, 36–7, 515
  - noise bandwidth, 38, 39, 149, 183, 212, 213, 254
  - noise temperature, 38
- Erlang B formula, 544
- ERMES *see* European Radio Messaging System (ERMES)
- error(s)
  - burst, 237, 422, 423, 428–30, 452, 453, 482, 484
  - control techniques, 355, 422–53
  - polynomial, 427, 428
  - propagation, 144, 148, 179, 236
  - quantization, 115, 469, 472
  - random, 422, 452, 482
  - squared, 109, 110, 118, 169, 405, 466
- error vector magnitude (EVM), 514, 522
- estimating filter, 114
- ETSI *see* European Telecommunications Standards Institute (ETSI)
- European Radio Messaging System (ERMES), 578, 579
- European Telecommunications Standards Institute (ETSI), 578, 582, 587
- EVM *see* error vector magnitude (EVM)
- Evolved Packet Core System Architecture Evolution, 597
- EX-OR phase detector, 534
- extrinsic information, 419, 432, 433
- eye
  - diagram, 128, 129, 132–4, 143, 147, 148, 150, 156, 270, 274, 287, 291, 293, 298–302, 306, 309, 317, 329
  - opening, 128, 132, 297, 298, 300, 302, 303
- fading
  - channel, 4, 203, 204, 207, 264–70, 322, 328, 335, 343, 355, 357, 366, 401, 410, 421, 423, 456, 467, 525, 561, 565
  - fast, 4, 189, 190, 193–200, 207, 236, 251, 272, 282, 299, 314, 322, 356, 367, 372, 401, 525, 532, 533, 569, 588
  - frequency-selective, 4, 200–204, 267–70, 272, 322, 326, 329, 335, 336, 343, 355, 357, 367, 401, 421, 547, 565, 583, 591
  - loss, 190
  - Rayleigh, 203, 206, 208–11, 265, 267, 269, 270, 274, 306, 310, 320, 321, 367, 373, 463–5, 559, 561, 568–70, 591

- fast
  - Fourier transform (FFT), 106–8, 329, 333, 334, 337, 339, 345, 418
  - frequency shift keying (FFSK), 283, 580
- FDM *see* frequency domain multiplexing
- FEC *see* forward error correction (FEC)
- feedback
  - circuit, 71–3, 143, 145, 178, 399, 402, 403, 411, 412, 471, 508, 518
  - gain, 508
- Feed-forward AGC, 528
- Felbart's little theorem, 182
- FFSK *see* fast frequency shift keying (FFSK)
- FFT *see* fast Fourier transform (FFT)
- FH *see* frequency hopping (FH)
- field programmable gate arrays (FPGA), 493
- finite impulse response (FIR), 65, 102
  - filter, 97–8, 101, 113
- first available, 554, 557
- flywheel effect, 177
- FOMA, 595
- forward error correction (FEC), 302, 422, 596
- Fourier
  - integral, 15, 17–21, 23, 79, 82–5, 126
  - series, 15–21, 80, 368
- Fourier transform, 17, 19–26, 103–5
  - discrete Fourier, 85, 88, 102–8, 337
- FPGA *see* field programmable gate arrays (FPGA)
- FPLMTS, 481, 592
- framed signal, 2, 3, 176, 177
- free distance, 454, 455
- frequency
  - division multiplexing, 182, 183
  - hopping, 322–4, 327, 328, 423, 565, 566, 585, 601, 605, 608
  - multiplication, 221, 233, 234, 236
  - sampling theorem, 77–8, 105
  - selective fading, 4, 200–204, 267–70, 272, 322, 326, 329, 335, 336, 343, 355, 357, 367, 401, 421, 547, 565, 583, 591
  - synthesizer, 495, 501, 502
- frequency shift keying (FSK), 220, 226–8, 246, 265, 267, 283, 296, 322, 360, 504–7, 535, 536, 577, 579, 580, 584, 605
- Friis formula, 189
- G.729, 479–81
- Gallager, R., 444
- Gaussian
  - distribution, 45, 53, 183, 450
  - filtered MSK, 299–302
  - noise, 41, 44, 48, 51, 150, 151, 156, 183, 237, 242, 257, 258, 260, 263, 270, 272, 276–7, 335, 347, 360–364, 434, 435, 446, 454
- general packet service (GPRS), 592
- generalized
  - Hamming window, 102, 504
  - partial response, 142, 298
  - tamed frequency modulation (TFM), 298–9
- generator matrix, 425, 445
- ghost, 73, 401
- GI *see* guard interval (GI)
- Glavieux, A., 430
- GMSK *see* Gaussian filtered MSK
- group
  - delay, 57, 69
  - demodulator, 535, 537
  - modulator, 535, 537
- Group Special Mobile (GSM), 583–8, 592, 595
- guard interval (GI), 330–334
- half-duplex, 501, 599
- hamming distance, 424, 430
- hand
  - off, 544, 548–50, 582, 587, 590, 596
  - over, 548, 549, 600, 606
- H-ARQ, 422, 452, 453
- Hata, M., 191
- HDLC, 453, 581
- HDR *see* high data rate (HDR)
- heterodyne receiver, 495–7
- high
  - data rate, 329, 550, 559, 565, 579, 595, 596
  - Q resonator, 502
- high data rate system, 559, 595, 596
- Hilbert transform, 35, 70, 340, 499
- Hirade, K., 269, 299
- home location register (HLR), 550, 587, 588
- homodyne, 497
- H(hybrid)PSK/HPSK, 521, 594



- HSCSD, 592
- HSDPA, 596, 597
- HS-DSCH, 596
- HSPA+, 597
- HSUPA, 596–8
- hunting mode, 177
- hybrid
  - ARQ, 567, 596, 598
  - channel assignment, 554
- IDFT *see* inverse discrete Fourier transform (IDFT)
- IEEE
  - 802.11, 600–605
  - 802.11a, 600, 604
  - 802.11g, 600, 604
  - 802.16, 599, 600, 606
  - 802.16m, 600
  - 802 committee, 599, 600
- IIR *see* infinite impulse response (IIR)
- Ikoma, M., 306
- image
  - frequency, 70, 495–8
  - rejection, 497–500
- impulse response, 2, 31, 43, 56–60, 62, 63, 65–70, 72, 73, 77, 85–7
- IMT
  - 2000, 481, 592, 595
  - DS, 593
  - FT, 593
  - MC, 593, 594
  - SC, 593, 595
  - TC, 593, 594
- infinite impulse response (IIR), 65, 98
- integrate-and-dump filter, 69, 156, 254, 257, 262, 263, 318, 367, 368, 528
- integrated services digital network (ISDN), 588, 597
- inter-cell (co-channel) interference, 566, 572, 573, 575
- interleaver, 430, 439, 442, 444
- intermodulation, 494, 495, 557
- intersymbol interference, 4, 128, 129, 145, 160–162, 251, 255, 260, 263, 264, 267, 292, 297, 300, 306, 317, 319, 326, 328, 330, 332–5, 399, 405, 408, 410–412, 417, 419–22, 467, 565, 598, 603
- inverse
  - circuit, 72–5, 101, 102, 113, 399, 411, 420, 514, 518, 519
  - matrix lemma, 121
  - matrix method, 388–9
  - modulation, 233, 235, 236
- inverse discrete Fourier transform (IDFT), 104, 106, 108, 339, 343–5
- irreducible error, 233, 266–9, 355, 364–6
- IS
  - 95, 590, 592, 594
  - 127, 481
  - 136, 592, 595
  - 95A, 592
  - 95B, 592
  - 641EFR, 592
- ISM band, 601, 605
- iterative turbo decoding, 441–2
- ITU
  - R, 599
  - T, 479
- Jager, F., 281, 294
- Jamming, 324, 456
- Japanese digital cellular system, 479, 493, 494, 591
- Jitter, 101, 175, 176, 233
- Kalman algorithm, 411, 421
- Kanai, 546, 557
- Karhunen-Loève series expansion, 165
- Kirchhoff-Huygens approximation, 214
- 625k-MC mode, 608
- LDPC *see* low density parity check (LDPC)
- least mean square (LMS), 118, 119, 406–9, 411, 457, 458, 469, 520
  - algorithm, 118, 408, 458, 520
  - method, 109–11, 118
- line
  - coding, 32, 127, 137–48
  - spectra, 77, 337
- linear
  - modulation, 36, 220–221, 231–3, 271, 275, 276, 281, 313–22, 507, 564, 585, 588
  - phase circuit, 98
  - system, 36, 55–75, 92, 94, 106, 112, 114, 115, 221, 388

- linear predictive coding (LPC)
  - predictive parameters, 477
  - vocoder, 482
- linearity, 21, 56–8, 64, 81, 86, 510, 522, 523, 525
- LLR *see* log likelihood ratio (LLR)
- LMS *see* least mean square (LMS)
- location registration, 544, 548, 550–551
- log likelihood ratio (LLR), 153, 155, 434–6, 450
- long-term evolution (LTE), 322, 345, 453, 526, 549, 565, 567, 597, 598
  - LTE-Advanced, 599
- low density parity check (LDPC), 155, 183, 430, 444, 445, 447, 449, 599
- M* algorithm, 398
- Macrocell, 548–9, 556, 557, 599
- MAP decoding, 446
- M*-ary PSK, 224, 242–3, 246, 250
- Maseng, T., 305
- matched filter, 69, 139, 150, 151, 157–61, 169–71, 174, 176, 237–9, 241, 244–6, 248, 267, 288, 292, 323–6, 341, 342, 410, 412, 420, 421, 435, 466, 467
- maximal throughput, 559
- maximum a-posteriori (MAP) algorithm, 436–41
- maximum likelihood
  - detection, 397
  - receiver, 170
  - sequence estimation, 299, 397, 538
- Max-Log-MAP Algorithm, 442–4
- MCA, 577
- media access control (MAC), 600
- memory effect, 515–18
- merging, 174, 455
- metric, 172, 173, 270, 430, 443
- microcell, 548–9, 554, 556, 557
- MIMO *see* multi-input multi-output (MIMO)
- minimum distance, 128, 137, 193, 424, 428, 429, 454
- minimum mean square
  - error, 275, 396, 420, 459, 461, 463, 465, 467, 519, 520
- minimum shift keying (MSK), 141, 270, 282–94, 296, 297, 300, 303, 305, 505, 532
- minimum-weight vector, 426, 428
- MLD *see* maximum likelihood detection (MLD)
- MMSE *see* minimum mean square error (MMSE)
- mobile
  - assisted hand-off, 549–50
  - radio channel, 189–217
  - station, 494–5
- mobile broadband wireless access (MBWA), 608
- MOBITEX, 580
- mod  $2\pi$  decision, 318
- modulation(s)
  - amplitude, 22, 34, 220, 340
  - constant envelope, 220–221, 275, 282–314, 525, 527, 605
  - continuous phase, 305
  - index, 226–8, 231, 232, 253, 256, 262–4, 267, 282, 283, 287, 294, 297, 299, 302, 305–8, 310–313, 366, 368, 369, 504–6, 535
  - inverse, 233, 235, 236
  - linear, 36, 220–221, 231–3, 271, 275, 276, 281, 313–22, 507, 564, 585, 588
  - phase, 34, 220, 304
  - quadrature amplitude, 35, 36, 220, 229, 320, 329, 343, 347, 348, 395, 412, 456, 569, 584
  - singe-side-band, 35, 70, 340, 507
  - trellis-coded, 453–6, 466
- modulo operation, 74, 181, 399, 400, 411, 412
- moving average circuit, 97, 530
- M*-sequence, 270, 328
- MSK *see* minimum shift keying (MSK)
- Multi
  - input multi-output, 7, 183, 355, 375–401, 412, 418, 568, 597, 604
  - pulse coding, 476–7, 480
  - user MIMO, 398–401
- multicarrier transmission system, 329–43, 535
- multi-input multi-output (MIMO), 7, 183, 355, 375–401, 412, 418, 568, 597, 604
- multiple access, 7, 8, 182–3, 305, 325, 543, 551–3, 563–6, 580, 585, 589–91, 593, 595, 601
- Murota, K., 299

- Nagata, Y., 314, 513
- near-far problem, 204–5, 292, 566
- negative feedback, 72, 234, 314, 315, 403, 508
- NMT, 577
- noise
  - band-pass, 44–9
  - bandwidth, 38, 39, 149, 183, 213, 254, 262, 570
  - blanking, 482
  - click, 257, 259, 263, 370, 482, 483
  - equivalent, 38, 39, 149, 183, 212, 254
  - figure, 38, 39, 149, 212, 213, 493, 568, 570
  - Gaussian, 41, 44, 48, 51, 150, 151, 156, 183, 237, 242, 257, 258, 260, 263, 270, 272, 276–7, 335, 347, 360–364, 434, 435, 446, 454
  - pseudo-, 178, 270, 323, 590
- nonlinear
  - circuit, 73, 75, 251, 266, 275–6, 514
  - distortion, 72, 270, 271, 313, 314, 335, 507, 508, 511, 514, 516, 517, 519, 521, 523
- nonlinearity compensation
  - methods, 507, 525
  - technique, 494
- non-return-to-zero (NRZ)
  - signal, 135, 156, 224, 292
  - signaling, 156, 157, 174, 224, 226, 228, 254
  - signalling system, 161
  - waveform, 31, 32
- NTT Docomo, 592, 595
- Nyquist
  - filtered multilevel FM, 306
  - I criterion, 44, 161, 162, 260, 349
  - I pulse waveform, 503
  - I signaling system, 157, 164
  - II, 134
  - III, 293, 294, 296–8, 300, 311
- Nyquist's
  - first criterion, 128–32
  - second criterion, 132–4
  - third criterion, 134–5
- OFDM *see* Orthogonal frequency division multiplexing (OFDM)
- offset QPSK, 223–5, 285–7, 316, 521
- Okumura, 191
- omni cell, 545
- Open System Interconnection (OSI), 581
- open-loop control, 513
- optimum receiver, 164–71, 174–5, 297, 410, 419
- orthogonal
  - frequency division multiplexing, 29, 329–37, 339, 346–9, 418, 420, 521, 522, 565, 598–600, 603, 604, 606, 607
  - series expansion, 520–521
  - signals, 26, 28–31, 182, 183, 283
- orthogonality, 21, 29, 31, 80, 327, 329, 341, 386, 565, 566, 596
- orthonormal functions, 165, 166, 169
- out-of-band frequency components, 519, 522
- out-of-sight communication, 189
- overlaid system, 548–9, 556
- packet
  - reservation multiple access, 558–9
  - switched system, 552
- paging, 5, 281, 544, 548, 550–551, 587
- systems, 5, 367, 578–9, 582
- PAPR *see* peak to average power ratio (PAPR)
- parity-check matrix, 426, 444, 446, 449
- Parseval's equation, 26
- partial response
  - code, 142, 147
  - generalized, 142, 298
- partial transmit sequence, 521
- path
  - loss, 190–192, 209, 210, 212, 213, 568
  - metric, 173
- PCM *see* pulse code modulation
- PD *see* predistorter (PD)
- PDC *see* personal digital cellular (PDC)
- peak distortion
  - criteria, 407–9
  - to-average power ratio, 183, 334, 335, 345, 514, 515, 521–2
- peak to average power ratio (PAPR), 183, 334, 335, 345, 514, 515, 521–2
- perceptual weighting factor, 476
- performance criteria, 405–9
- periodic function, 15, 16, 19–21, 77–80, 84, 103, 104
- personal communication, 587

- personal digital cellular (PDC), 481, 591, 592
- Personal Handy System (PHS), 582–3
- perturbation method, 116, 519
- PF *see* proportional fairness (PF)
- PFS *see* proportional fair sharing (PFS)
- phase
  - ambiguity, 224, 236, 287, 289
  - detector, 503, 509, 528, 530–533
  - locked loop, 233, 306, 501
  - modulation, 34, 220, 304
  - shift keying, 220–225, 305, 320
  - shift method, 499
- phase detector, 503, 509, 528, 530–533
  - quadrature, 530, 531
- $\pi/4$  shifted QPSK, 224–6, 228, 241, 273, 306, 310, 313–21, 364, 366, 367, 494, 508, 509, 533, 534, 582, 588, 589, 591
- $\pi/2$  Shifted BPSK, 222–5, 228, 283, 285, 286, 306
- picocell, 548, 549, 599
- pilot signal(s), 236, 322, 334, 528, 593, 595, 597
- pitch, 473, 475, 479, 480
  - parameters, 477
- PLL *see* phase-locked loop
- PLL-QPSK, 306–10, 312, 314
- PN *see* pseudo noise (PN)
- POCSAG *see* Post Office Code Standardization Advisory Group (POCSAG)
- Poisson arrival rates, 608–9
- polar loop method, 509
- polling system, 551, 580
- Post Office Code Standardization Advisory Group (POCSAG), 578
- posteriori probability, 153, 167, 168, 436, 437, 446
- power
  - delay profile, 200, 202, 203
  - transfer function, 38, 56, 272
- power spectral density (PSD)
  - of filtered noise, 43
  - of noise, 42–3
  - passband noise, 45–8
  - random digital signal, 229–32
- preamble signal, 233, 371, 565
- precoding, 74, 144, 145, 297, 399–401, 411–18
- prediction, 191, 375, 409, 472, 525
  - algorithm, 411
- predistorter (PD), 512–16, 518, 526
- pre-equalization, 411–18, 521
- preferred pair, 328
- prescaler, 502
- press-to-talk, 501, 577, 579
- PRMA *see* packet reservation multiple access
- probability
  - density function, 12, 40, 44, 45, 48–54, 151, 156, 193, 195, 206, 207, 209, 216, 242, 245, 248, 253, 255, 266, 357–62, 364, 369, 372, 373, 450, 465, 469, 485–6, 570, 571, 575
  - distribution function, 39, 52, 206, 209, 363, 571
- proportional
  - fair scheduling, 112, 563
  - fair sharing, 559, 561–3, 596
  - fairness, 559–63
- protection ratio, 206, 207, 544, 545, 590
- protocol, 3, 581, 587, 600, 601
- pseudo noise (PN), 178, 270, 323, 590
  - sequences, 323
- PSI-CELP, 479
- PSK *see* phase shift keying
- PSTN *see* public switched telephone network
- public
  - key cryptosystem, 180–182
  - switched telephone network, 580, 581, 588
- pulse
  - code modulation, 115, 141, 470, 482, 592
  - duty, 135, 137, 326
  - shaping, 127–37, 141, 142, 174
- $Q$ -value, 175, 233
  - of the resonator, 175
- QAM *see* quadrature amplitude modulation (QAM)
- QCELP, 481, 590
- QPSK *see* quadrature phase shift keying
- QR decomposition, 397
- quadrature
  - AM, 35, 36, 220, 229, 320, 329, 343, 347, 348, 395, 412, 456, 569, 584
  - phase shift keying, 7
- quadrature amplitude modulation (QAM), 35, 36, 220, 229, 320, 329, 343, 347, 348, 395, 412, 456, 569, 584
- Qualcomm Inc., 559, 590, 595
- quantization, 75, 115, 471, 472, 475, 477, 479, 481
  - noise, 469

- quarter wavelength line, 538–9
- raised cosine
  - roll-off filter, 131–4, 367
  - roll-off Nyquist-I filter, 317–19
  - waveform, 136
- Raith, K., 421
- RAKE receiver, 160, 325–8, 357
- random
  - access system, 551–3
  - frequency modulation (FM), 198–200, 233, 236, 266–7, 270, 355, 364–6
- Rayleigh
  - distribution, 44, 45, 195, 200, 265, 357, 367, 373, 576
  - fading, 203, 206, 208–11, 265, 267, 269, 270, 274, 306, 310, 320, 321, 367, 373, 463–5, 559, 561, 568–70, 591
- RCR *see* Center for Radio Systems (RCR)
- received signal strength indicator (RSSI), 528, 529
- recursive least square (RLS), 118–20, 123, 411, 466
- regeneration of carrier signal, 31, 233
- reliability information, 432, 442, 444
- re-modulation, 233, 235, 236
- repeater station, 566
- reuse partitioning, 547, 576
  - self-organized, 557
- Rice, S.O., 248, 253
- RLS *see* recursive least square (RLS)
- roll-off factor, 131–4, 140–142, 147, 148, 262, 263, 273, 293, 295, 298, 307, 309, 311, 313, 314, 317, 318, 329, 347, 503
- round-robin (RR), 397, 559, 561–3
- RPE-LPC/LTP, 476
- RR *see* round-robin (RR)
- RSA cryptosystem, 180, 181
- RSSI *see* received signal strength indicator (RSSI)
- Saegusa, N., 306
- sampling
  - down-sampling, 98–101
  - sub-sampling, 100–101
  - up-sampling, 98–101
- sampling theorem, 75–8, 88, 95, 98, 100, 104, 105, 271, 334, 337, 340
  - frequency, 77–8, 105
- Schwartz inequality, 158, 163
- scrambling
  - self-synchronized, 178, 179
  - synchronized, 1, 179
- SDM(A), 183
- sector cell, 545–6, 569, 572, 575
- selection (switching) combining, 356–9, 364
- selective mapping, 521
- Selective Repeat, 452, 453
- self-organized reuse partitioning (SORP), 557
- self-synchronized system, 178, 179
- shadowing, 190, 193, 206–9, 212, 216, 217, 356, 367, 545, 568, 571, 572, 590
- Shannon, C.E., 183
  - limit, 430, 449, 450
- side information, 475, 521, 522
- signal(s)
  - analysis, 9–37
  - carrier, 2, 3, 31, 34, 35, 139, 141, 194, 220, 221, 226, 233–5, 237, 282, 289, 291, 322, 345, 418, 497, 533, 564, 590, 591
  - complex baseband, 220, 230, 517
  - digital, 1, 3, 31–4, 36, 99, 112, 127, 141, 149, 164, 170, 176, 177, 182, 219–21, 224, 226, 228, 231, 282, 283, 329, 337–41, 343–5, 347, 370, 397, 401, 402, 419, 420, 434, 493, 494, 502, 503, 506, 507, 514, 529, 537, 538, 565, 566, 577
  - flow graph, 65, 93
  - to interference plus noise, 459–66
  - to interference ratio, 547
  - modulated, 2, 34–6, 135, 141, 149, 175, 194, 219–21, 224, 230, 233–6, 245, 257, 271, 272, 275–6, 282, 293, 305, 306, 320, 322, 323, 329, 343, 346, 356, 454, 495, 497, 498, 514, 528, 533, 564
  - to noise power ratio, 37, 149, 151, 157, 158, 162, 163, 176, 183, 184, 213, 233, 238, 241, 248, 254, 257–9, 261, 263, 264, 291, 324, 328, 357, 366, 375, 377–9, 385, 390, 392–4, 411, 414, 418, 420, 421, 450, 465, 470, 483, 521, 559
  - orthogonal, 26, 28–31, 182, 183, 283

- signal(s) (*cont'd*)
  - pilot, 236, 322, 334, 528, 593, 595, 597
  - postamble, 371, 372
  - prediction, 114, 115
  - zero IF, 271
- SIMO *see* single-input multi-output
- single
  - carrier frequency-division modulation, 343–9
  - input multi-output, 375
  - side-band modulation (SSB), 340
- slow FH, 323
  - system, 327, 328, 566
- soft-output Viterbi algorithm (SOVA), 442, 443
- software simulation, 274
- SORP *see* self-organized reuse partitioning (SORP)
- space
  - division multiplexing, 375, 390, 418, 599
  - time block code, 385
  - time trellis code, 385
- spatial correlation, 198, 199
- spectral density
  - energy, 26–8
  - power (*see* power spectral density)
- spectrum efficiency
  - of a cellular system, 302, 544
  - spreading (SS) codes, 467, 565
- split-Phase code, 139–40
- spread spectrum system(s), 160, 322–8, 590
- stability of a system, 62, 91
- standing waves, 190, 193
- statistically multiplexing, 559
- STBC *see* space-time block code
- steepest decent, 118, 519
  - method of, 117
- Stjernvall, J.-E., 421
- STTC *see* space-time trellis code
- subcarrier system, 282, 578
- subkeys, 179
- superheterodyne, 495–500
- surviving path, 173
- symbol timing
  - recovery, 175–6
  - regeneration, 176
- symbol-by-symbol decision, 174, 297
- SYNCH word, 176–8
- synchronization, 175–9, 235–7, 328, 333, 356, 419, 483, 565, 601
  - frame, 176–8
- syndrome, 426, 428, 446–8
- system
  - cost, 7–8, 581
  - function, 56, 88, 90, 92, 93
- table look-up method, 275, 503
- Takase, I., 306
- tamed frequency modulation (TFM), 294, 296–8, 300, 303, 304, 306, 314
  - generalized, 298, 299, 584
- tank circuit, 175, 233
- Tanner graph, 447
- TDD *see* Time division duplex
- Telecommunications Industry Association (TIA), 587
- teleterminal system, 580
- 3rd generation partnership project (3GPP), 481, 592, 593, 595, 597, 599
  - 3GPP2, 481, 593, 595, 599
- Thitimajashima, 430
- 3G system, 592, 593, 595–9
- throughput, 275, 345, 454, 552, 553, 559, 561, 563, 564, 567, 569, 573, 575, 576, 596
- time
  - division duplex, 501, 525, 567, 594, 597, 598, 608
  - invariant system, 55, 58, 64, 66, 85–7
  - resolution, 325
  - variant system, 58
- time division duplex (TDD), 501, 525, 567, 594, 597, 598, 599, 608
- time division multiplexing (TDM), 1, 8, 182, 183
- token passing system, 551
- Tomlinson–Harashima pre-coding, 399–401, 411, 412
- transfer function, 2–4, 56–8, 61–3, 65, 66, 69–72, 87, 88, 91, 94–7, 101, 113, 118, 128, 130, 131, 134–6, 139, 140, 142
- transmission control protocol/Internet protocol (TCP/IP), 581
- transmission delay, 552, 553, 592, 598
- transmit and receive duplexing, 501, 581, 585, 597, 600

- transmit power control, 205, 525, 566, 590, 595
- transversal filter, 97, 114, 118, 324, 402, 404
- trellis
  - coded modulation, 453–6, 466
  - diagram, 448, 455, 456
- turbo code(s), 155, 183, 430, 432, 433, 444, 445, 447, 451, 569, 592, 594, 596, 599
- 2G system, 595
- two-path model, 200, 202, 203
- two-way mobile radio, 579
  
- Uddenfeldt, J., 421
- ultra-wideband (UWB), 605, 606
- unipolar code, 137
  
- value-added networks, 581
- VCO *see* voltage controlled oscillator (VCO)
- vector coding, 411–13, 420
- vector sum excited LP (VSELP), 478, 479, 494, 588, 589
- Viterbi algorithm, 170–174, 410, 430, 443, 456, 466, 538
- vocal tract, 473, 474
- voice
  - coding, 1, 7, 114, 296, 355, 469–84, 564, 567, 579, 581, 582, 584, 585, 587–92, 594
  - operated transmission, 484, 494
  - scrambling, 5
- voice operated transmission (VOX), 484, 494
- voice over Internet protocol (VoIP), 567, 597
- voltage controlled oscillator (VCO), 234, 235, 501–4
- Volterra
  - power series, 517
  - series, 37
- VOX *see* voice operated transmission (VOX)
- VSELP *see* vector sum excited LP (VSELP)
  
- Walsh function, 29, 79, 565, 590
- wave propagation, 190, 274, 388, 462, 463, 545, 548, 553, 554, 567
- W-CDMA, 591, 593–7
- WiMAX
  - 2, 599, 600
  - 2+, 600
- window
  - Hanning, 102
  - rectangular, 102, 504
- Wireless Local Area Network, 337, 600–608
  
- $z$ -transform, 80–82, 84, 85, 88, 90–93, 142, 144, 148, 340, 402, 411
  - bilinear  $z$ -transform, 96, 97
- Zero forcing algorithm, 396
- Zero IF, 271, 497
- ZigBee, 606